

# Sélection d'objets vidéo basée sur une approche active de propagation de gribouillis

MARWEN NOURI<sup>1,2</sup>, EMMANUEL MARILY<sup>2</sup>, NICOLE VINCENT<sup>1</sup>

<sup>1</sup> LIPADE, Université Paris Descartes  
45 rue des saints pères, 75006 Paris, France

<sup>2</sup> Alcatel-lucent Bell Labs France  
Route de villejust, 91620 Nozay, France

<sup>1,2</sup>Marwen.Nouri@alcatel-lucent.com, <sup>2</sup>Emmanuel.Marily@alcatel-lucent.com, <sup>1</sup>Nicole.Vincent@mi.parisdescartes.fr

Résumé - Récemment, plusieurs approches innovantes sont apparues traitant le problème de la segmentation d'objets de haut niveau sémantique. La plupart de ces approches utilise l'interaction avec l'utilisateur, via le dessin de quelques gribouillis par exemple, afin de surmonter le gap sémantique. Dans le cadre de la segmentation d'objets vidéo, de telles interventions ne peuvent pas être réalisées sur chaque image de la vidéo. Un support automatique à l'utilisateur est nécessaire pour réduire ces efforts rendre ce type de segmentation possible. Dans cet article, nous présentons un nouveau système semi-automatique de sélection d'objets vidéo. Cette sélection est assurée par la modélisation et la propagation des gribouillis sous forme de courbe active. Les gribouillis sont propagés de manière cohérente de la première image au reste de la vidéo en désignant le même objet, initialement désigné par l'utilisateur, par un processus d'optimisation spatiotemporel.

Abstract - While the problem of motion tracking has been widely studied, tracking meaningful object still a complex issue. User interaction is needed to avoid a semantic gap. Nevertheless video object tracking implies excessive user interactions. In this paper, a new framework for interactive video object selection is described. This framework is built upon active curve based scribbles propagation. In this approach, the user can designate an object in a video by drawing scribbles in one video frame. These scribbles are propagated automatically to designate the same object in the next frames. The active curve model is adapted and new forces are created to govern the curves evolutions. Curves are tracked frame by frame. A spatiotemporal optimization is used to ensure a coherent propagation.

## 1 Introduction

L'analyse humaine du mouvement est très difficile à reproduire par ordinateur. La diversité des informations dont nous disposons, nous permet d'interpréter les mouvements dans les situations complexes (problème d'ouverture, confusion temporelle...). Dans le cadre de la vision par ordinateur, l'analyse du mouvement est une tâche primordiale pour la plupart des algorithmes de traitement de vidéo. Le suivi d'objets est un sujet ayant donné lieu à une littérature abondante. Il constitue un des axes majeurs des recherches dans le domaine. Une méthode de suivi générique n'existe pas, mais plusieurs approches ont pu démontrer leurs capacités d'adaptation à diverses problématiques. Nous pouvons citer l'approche proposée par Lucas et Kanade [11] qui est une approche de suivi ponctuelle; le mouvement est traité localement au niveau de points d'intérêts. Cette approche est utilisée et a inspiré plusieurs algorithmes récents. Aussi, on peut citer la méthode des contours actifs [7], qui est une approche plus globale se basant sur les silhouettes. Le modèle des contours actifs a connu un grand succès dans le domaine de la segmentation et a été adapté dans [8] pour le suivi d'objets déformables. La plupart des approches se basant sur les contours actifs définissent un ensemble de forces, permettant de faire tendre un contour initial vers la position qui délimite l'objet d'intérêt. Cette nouvelle

position correspond naturellement aux points de contour de l'objet. La force principale va donc guider le contour initial vers les points de plus fort gradient. Ce modèle a aussi introduit la notion de la possibilité d'impliquer l'utilisateur lors de l'initialisation. Ceci nous oriente vers les approches dites interactives. Les informations fournies par l'utilisateur permettent de surmonter le gap sémantique et d'extraire des objets de haut niveau plutôt des régions partageant un certain critère de similarité, par exemple. Dans le cadre du *digital matting* [1], il suffit à l'utilisateur de marquer un objet par un ou plusieurs gribouillis (Figure 1, 5a, 6a) pour l'extraire de l'image. La tâche devient beaucoup plus complexe quand on parle de *matting* vidéo car le nombre d'interactions nécessaires peut être énorme s'il n'y a pas d'assistance automatique.

Dans cet article, nous proposons une méthode permettant de suivre un objet en mouvement en le désignant par propagation de gribouillis. Ces gribouillis peuvent reproduire le marquage, initialement fait par l'utilisateur sur la première image de la vidéo, sur le reste de la séquence. La section 2 décrit les problèmes relatifs à la propagation de gribouillis et introduit leur modélisation sous forme de courbes actives. La section 3 présente et discute des différentes forces pour assurer une propagation cohérente. Les résultats ainsi que les expérimentations sont présentés dans la section 4.

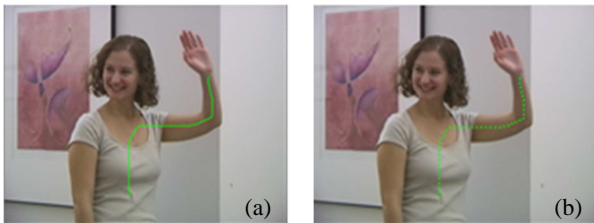


Figure 1: (a) Exemple d'un gribouillis dessiné par l'utilisateur. (b) Le gribouillis discrétisé en un ensemble de point.

## 2 Propagation de gribouillis

Nous proposons une nouvelle méthode ayant pour objectif de propager des gribouillis dessinés par l'utilisateur et désignant un objet sur une image de la vidéo au reste de la séquence. Dans cet objectif, nous avons choisi une modélisation des gribouillis sous la forme de courbes actives. En premier lieu, les caractéristiques et les problèmes relatifs à la propagation de courbes seront détaillés. Ensuite, notre approche et les démarche de résolution du problème sont présentées en tenant compte des différentes propriétés des courbes dessinées.

### 2.1 Problématique de la propagation de courbe

Pour suivre une courbe tracée manuellement sur une image, il s'agit de traquer un ensemble de points. Les méthodes classiques d'estimation de mouvement, telle que [11], ne permettent pas la propagation de ces points sur un nombre important d'images consécutives. Pour être correctement suivis, ces points doivent avoir un certain nombre de propriétés, par exemple ils doivent être des points d'intérêt [5]. En effet, image après image, les erreurs de suivi des points sont propagées et accumulées progressivement. Les méthodes de suivi ponctuelles sont très sensibles à la ressemblance de texture et au problème d'ouverture. En réalité les points de la courbe ne sont pas géométriquement indépendants. Les traiter séparément ne fera qu'ajouter de l'incohérence au suivi. On propose de traiter cet ensemble de points comme un tout, comme une courbe.

Quelle est la fonctionnalité de la courbe que l'on veut propager? Cette question se pose implicitement à l'utilisateur qui veut désigner des zones. La réponse à cette question aidera à déterminer les propriétés qui permettent de propager la courbe tout en lui conservant sa fonctionnalité. Nous nous plaçons dans l'hypothèse où l'utilisateur trace une courbe pour désigner un objet dans la scène. La courbe passe donc par un certain

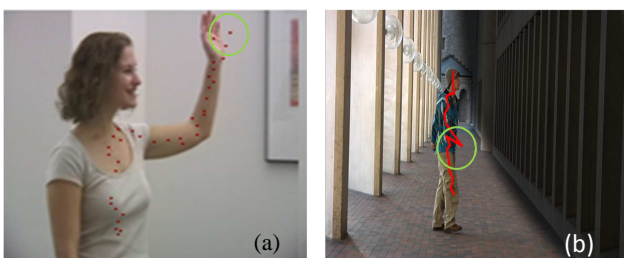


Figure 2: (a)(b) Résultats de la propagation basés sur le flux optique dans l'image 10 de la vidéo 'amira' et dans l'image 5 de la vidéo 'walking man'. Les erreurs sont indiquées par les cercles verts.

nombre de régions de différentes homogénéités qui composent l'objet. Elle est placée plutôt dans la partie médiane de ces zones comme on le voit sur la Figure 1. En effet la précision de l'utilisateur est faible, il est alors plus sûr pour lui de dessiner loin des bords.

Pour désigner le même objet dans les images suivantes de la vidéo, on souhaite que la courbe reste dans les zones considérées et se positionne le plus possible au milieu de celles-ci de manière à limiter la dérive. La courbe n'est pas caractérisée par un fort gradient. Il ne s'agit pas ici, comme c'est le cas dans de nombreuses études [7, 8] d'une courbe de contour qui limite un objet que l'on suit mais plutôt d'une courbe qui va définir les différentes parties d'un objet. Nous avons identifié trois contraintes permettant la propagation d'une courbe de tracée sur une image à l'instant  $t$  vers l'image figurant l'instant  $t+1$ : Si l'objet désigné se déplace, la courbe doit être déplacée en conséquence dans l'image. Il faut conserver les caractéristiques des zones traversées, en tenant néanmoins compte du fait une zone peut subir des changements aussi bien au niveau de la forme qu'au niveau de l'illumination. La courbe doit se déplacer vers les zones les plus homogènes, localement, pour minimiser les erreurs et essayer d'approcher la région médiane de chaque zone.

Dans notre cas, on ne peut pas avancer d'hypothèses sur la nature de la courbe dessinée par l'utilisateur. Les modèles des courbes de Bézier ou les Splines, par exemple, ne sont pas assez souples pour répondre aux trois contraintes exprimées ci-dessus.

Pour l'ensemble des régions composant l'objet, nous proposons de considérer la courbe globalement et de modéliser les forces qui conduisent à la repositionner au long de la vidéo. C'est une approche par courbe active qui est donc présentée.

Pour assurer une propagation cohérente, il est nécessaire d'introduire ces nouvelles contraintes dans la gestion de l'évolution de la courbe. Nous les avons transcrit ces contraintes sous forme de forces et défini leurs fonctionnelles d'énergies associées permettant de repositionner la courbe sur les images suivantes. L'estimation de la position de la courbe d'une image à l'autre se fait via un processus de minimisation d'énergie basé sur l'algorithme greedy [10] des contours actifs.

## 3 Modélisation des contraintes

Afin d'assurer une propagation cohérente de l'ensemble de points constituant un gribouillis désignant un objet composé de différentes régions, on a défini un ensemble de forces, les unes associées à la courbe elle-même et les autres liées aux données. Cet ensemble de forces correspond à l'intégration des différentes contraintes citées précédemment. L'objectif est de minimiser la somme des énergies correspondant aux différentes forces pour aboutir à une solution qui minimise les erreurs et qui est optimale pour l'ensemble de ces forces réunis.

### 3.1 Forces internes

Les forces internes sont là pour guider le comportement global de la courbe. Comme présenté dans [7], on utilise une force relative à la continuité et une autre relative à la courbure. L'énergie représentant la force de continuité va essayer de maintenir la cohésion des points et donc va uniformiser les distances séparant les couples de points successifs. L'énergie relative à la force de courbure va avoir une influence sur la rigidité de la courbe. Ces deux fonctionnelles d'énergie vont influencer la forme de la courbe au fur et à mesure de son évolution. Des poids  $\omega_1$  et  $\omega_2$  sont utilisés pour privilégier un comportement ou un autre. Pour chaque point  $p$  de la courbe, l'énergie interne s'écrit :  $E_{interne}(p) = \omega_1 E_{unif}(p) + \omega_2 E_{courb}(p)$  (1)

Si le gribouillis a pour objectif de suivre globalement un objet simple, la valeur des poids ne sera pas la même que si c'est un objet complexe qui est suivi. Nous proposons d'étudier la distribution des courbures dans l'image initiale pour fixer les poids.  $\omega_2$  est d'autant plus faible qu'il existe des points de forte courbure dans l'image précédente.

### 3.2 Forces externes

La définition des forces externes est plus sensible. Elle est directement reliée à la nature du problème. Pour la propagation de gribouillis, donc d'un ensemble de points désignant un objet de haut niveau, nous avons décidé de modéliser les contraintes définies précédemment par un ensemble de trois forces :

- Une relative à l'estimation du mouvement en chaque point de la courbe
- Une autre relative à la similarité des couleurs
- Une dernière qui met en évidence une direction à privilégier pour la propagation

Le mouvement d'un objet peut être la composition de différents mouvements des différentes parties qui le composent. Donc, estimer le **mouvement** localement au niveau des points est nécessaire pour assurer que la courbe puisse coller dans sa globalité au mouvement de l'objet. Notons  $p'_t$  l'image estimée par le flux optique [11] du point  $p_t$  de la courbe dans l'image  $t$ . Notre première force va tenter de minimiser la distance entre

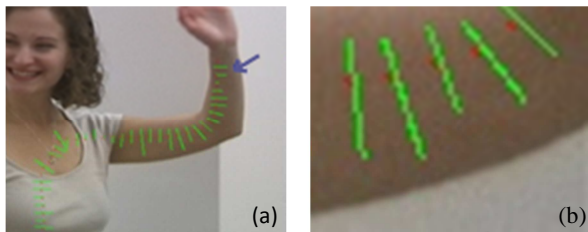


Figure 3: (a) segment de similarité calculé en chaque point de la courbe. (b) Zoom sur (a)

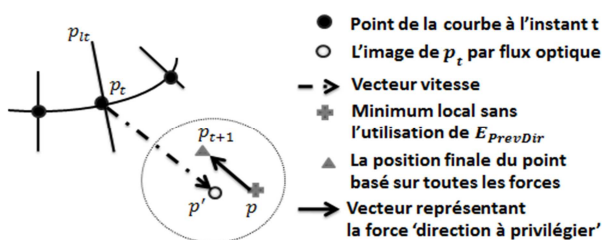


Figure 4: Représentation de la force direction à privilégier

$p$  et  $p'_t$ , pour obtenir la nouvelle position  $p_{t+1}$  dans l'image  $t+1$ . Ainsi on note sa fonctionnelle d'énergie:

$$E_{mvt}(p) = |p, p'_t|$$

Entre les images aux instants  $t$  et  $t+1$ , un point de la courbe doit rester dans la même zone dans l'objet. Une zone peut être caractérisée par sa couleur ou aussi par sa texture. La caractérisation par la texture est plus coûteuse en temps de traitement, on va alors limiter notre étude à une **couleur** moyenne sur un voisinage de chaque point. Nous avons utilisé le CIE Lab comme espace de couleur. Ce dernier, plus adapté que l'espace RVB, il nous permet la distinction entre les composantes chromatiques et la luminance. Cette deuxième force doit pousser alors un point  $p_t$  de la courbe à évoluer vers une nouvelle position  $p$ , dans l'image  $t+1$ , ayant des couleurs similaires en respectant une tolérance par rapport au changement d'illumination. L'énergie correspondante est calculée en se rapportant à une distance euclidienne pondérée comme suit :

$$E_{couleur}(p) = (cm_{pt}(a) - cm_p(a))^2 + (cm_{pt}(b) - cm_p(b))^2 + \frac{1}{4} * (cm_{pt}(l) - cm_p(l))^2 \quad (2)$$

La dernière force va essayer de reproduire le fait que l'utilisateur manque de précision lorsqu'il dessine et qu'il évite les bords (ce qu'il lui évite de sortir de l'objet). En effet cette force pousse la courbe à rester au milieu des différentes zones de l'objet. Cette force pousse les points vers les endroits de plus grande homogénéité. Pour cet effet, nous détectons les régions homogènes aux alentours de la courbe en nous basant sur la couleur de segments orthogonaux à la courbe en chaque point (Figure 3). Notre objectif, en partant de l'état d'un point à l'instant  $t$  est de **privilégier la propagation dans une direction** par rapport aux autres (Figure 4). En fait dans cet exemple c'est dans le sens du bras que l'on trouve la plus grande longueur homogène. Pour calculer l'énergie relative à cette force, on estime la normale à la courbe en chaque point  $p_t$ , ensuite on calcule sur cette droite un segment de similarité. Le segment associé à chaque point  $p_t$  est le segment maximum de couleur uniforme. Ce qui veut

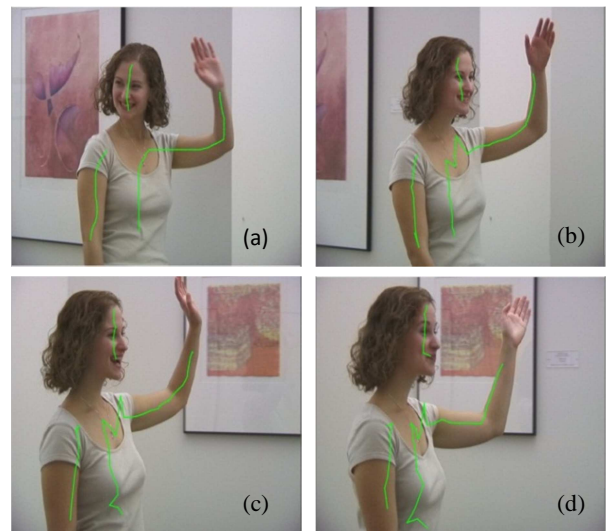


Figure 5: (a) Trois gribouillis dessinés par l'utilisateur sur l'image 1 de la séquence 'Amira'. (b)(c)(d) Respectivement, les résultats de la propagation à partir de (a) sur les images 8, 23 et 30.

dire que tous ces points ont une couleur similaire à  $p_t$ , et il est maximum au sens de la longueur. Les extrémités de ce segment sont les points limites de la tolérance couleur. On dénote par  $p_{lt}$  celui le plus éloigné de  $p_t$ . L'énergie correspondante s'écrit alors:

$$E_{privDir}(p) = |p, p'_{lt}| \quad (3)$$

Finalement, l'énergie externe va être composée de trois termes pondérés et on écrira l'énergie globale :

$$E_{globale}(C) = \sum_{p \in C} \left( \omega_1 E_{unif}(p) + \omega_2 E_{courb}(p) + \omega_3 E_{mvt}(p) + \omega_4 E_{couleur}(p) + \omega_5 E_{prevDir}(p) \right) \quad (4)$$

Les différents poids ont été déterminés expérimentalement, comme c'est le cas des algorithmes basés sur la minimisation d'énergie. Les résultats sont présentés dans la section suivante sur différentes vidéos tous en gardant le même ensemble de poids fixes.

## 4 Résultats expérimentaux

Notre système propose une interface utilisateur permettant de dessiner un ou plusieurs gribouillis sur importe quelle image de la vidéo. Ces gribouillis sont alors propagés en temps réel aux images suivantes de la vidéo. L'utilisateur peut intervenir à tout moment pour arrêter la propagation ou pour pointer une région de l'objet qui n'était pas visible au pare avant. Nos résultats sont mieux visibles sous forme vidéo mais nous suffirons de présenter quelques images dans cette section. Nous avons testé notre approche sur trois séquences standard de la littérature [1,2] (Figure 5, 6) (séquences de personnes qui marchent, filmé par une caméra mobile). Il est difficile de mesurer qualitativement le succès d'une approche tel que la nôtre. Une mesure possible peut être basée sur le nombre d'images dans lesquels l'objet initialement indiqué par l'utilisateur continu à être désigné par notre algorithme.

Pour comparer notre méthode on a implémenté un propagateur de gribouillis basé sur le flux optique, qu'on notera OFBP. Les figures 5a et 6a montrent deux exemples de gribouillis en entrées par l'utilisateur. Nos résultats sont présentés et comparés à l'implémentation OFBP dans le tableau Tab1. La Figure 2 montre que même dans des cas assez simple la propagation échoue assez rapidement en se basant que sur le flux optique.

**Tab 1: Nombre d'images dans lesquels l'objet initialement choisit continu à être désigné**

Méthode \ Vidéo	Amira	Adam Lib (29 images)	Walking man
OFBP	11	26	5
Notre méthode	30	29	25

Les différentes contraintes gérées par notre méthode rendent la propagation plus robuste. Notre système pourra être encore amélioré surtout pour la gestion des cas de changement brusque d'illumination combiné à des erreurs d'estimation de mouvement, comme c'est le cas dans l'image 25 de la séquence 'walking man' Figure 6d.

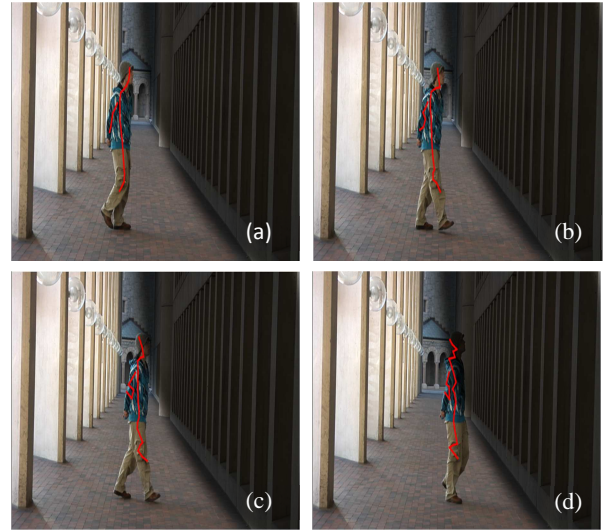


Figure 6: (a) l'utilisateur désigne la personne par deux gribouillis dans la 1<sup>er</sup> image de la séquence 'walking man'. (b)(c)(d) Respectivement, les résultats de la propagation à partir de (a) sur les images 7, 14 et 25.

## 5 Conclusion

Dans cet article, nous avons présenté une nouvelle approche pour la propagation de courbes dessinées manuellement. Cette approche consiste à la formulation du problème de désignation d'objet vidéo par des gribouillis sous la forme d'un problème d'évolution de courbe active. Cette évolution est assurée par un mécanisme de minimisation d'énergies. Nous avons aussi défini l'ensemble de forces nécessaires pour garantir une propagation cohérente tout en continuant à pointer l'objet initialement désigné par l'utilisateur.

## 6 Références

- [1] Jue Wang, Pravin Bhat, R. Alex Colburn, Maneesh Agrawala, and Michael F. Cohen, "Interactive video cutout," *ACM Trans. Graph.*, vol. 24,, pp. 585–594, July 2005.
- [2] Xue Bai, Jue Wang, David Simons, and Guillermo Sapiro, "Video snapcut: robust video object cutout using localized classifiers," *ACM Trans. Graph.*, vol. 28, no. 3, 2009
- [3] A. Levin, D. Lischinski, and Y. Weiss, "A closed-form solution to natural image matting," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 228–242, Feb. 2008.
- [4] Michal Irani, "Multi-frame correspondence estimation using subspace constraints," *International Journal of Computer Vision*, vol. 48, no. 3, pp. 173–194, 2002.
- [5] Jianbo Shi and Carlo Tomasi, "Good features to track," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94)*, Seattle, June 1994.
- [6] P. Sand and S. Teller, "Particle video: Long-range motion estimation using point trajectories," *International Journal of Computer Vision*, vol. 80, no. 1, Oct. 2008.
- [7] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.
- [8] S. Lefevre and N. Vincent, "Real time multiple object tracking based on active contours," in *ICIAR, 2004*, pp. II: 606–613.
- [9] Patrick Etymgier, Florent S'egonne, and Renaud Keriven, "Active-contour-based image segmentation using machine learning techniques," pp. 891–899.
- [10] D. Williams and M. Shah, "A fast algorithm for active contours and curvature estimation," *CVGIP:IU*, vol. 55, no. 1, pp. 14–26, Jan. 1992.
- [11] S. Baker and I. Matthews, "Lucas-kanade 20 years on unifying framework," *International Journal of Computer Vision*, vol. 56, no. 3, pp. 221–255, Feb. 2004..