Un outil de segmentation pour la manipulation de séquences vidéo

Vincent Lepetit et Marie-Odile Berger LORIA/ INRIA Lorraine BP 239, 54506 Vandoeuvre les Nancy e-mail:lepetit@loria.fr, berger@loria.fr

Résumé

De nombreuses activités liées à la post production ou à la réalité augmentée visent à créer des images ou des séquences d'images à partir d'images réelles en ajoutant, en enlevant ou en modifiant des objets dans la scène. Il est donc indispensable de disposer de méthodes permettant de détourer précisément un objet dans une séquence d'images. Nous proposons ici une méthode semi-interactive très simple d'utilisation permettant d'effectuer cette tâche: une fois que l'utilisateur a détouré l'objet dans un certain nombre de vues clés, notre méthode permet la détermination automatique et précise de l'objet dans les vues intermédiaires en utilisant à la fois des techniques de reconstruction 3D et de suivi. Les points de vue, ainsi que l'incertitude associée, étant calculés automatiquement pour chaque image de la séquence, nous pouvons obtenir une reconstruction locale de l'objet grâce aux détourages effectués dans les images clés. Ceci permet d'obtenir une prédiction fiable de l'objet, associée à une région d'incertitude, et d'affiner son détourage par des méthodes de suivi en tenant compte de la région d'incertitude. De nombreux exemples illustrent notre méthode. Des exemples complets montrant les vidéos des résultats sont visibles à l'URL:

http://www.loria.fr/~lepetit/Occlusions.

Mots Clef

détourage, reconstruction, suivi, point de vue, incertitudes.

Abstract

The main goals of image editing systems are to transform, to edit or manipulate one or more images to produce desired results. Tasks such as accurate selection, warping ... are often tedious and time consuming. Thus, this article focus on a computer vision technique that alleviate considerably the efforts involved in editing images and video. In this paper, we propose a semi automatic method for accurate object segmentation which makes use of key-views. Once the user has outlined the object in the key-views, our system detects automatically this object in the intermediate views. Our method makes use of both 3D reconstruction and tracking-based methods: a region of interest that contains the occluding objects is first computed from the outlined silhouettes. One of the main contribution of this paper is that this approach takes into account the uncertainty on the computed interframe motion. Then a deformable region-based approach is used to recover the actual occluding boundary from this prediction within the region of interest. Results are presented demonstrating object extraction in various examples. Videos can be shown at URL

http://www.loria.fr/~lepetit/Occlusions.

Correspondance: Marie-Odile Berger LORIA-INRIA Lorraine BP 239 54506 Vandoeuvre les Nancy cedex

1 Introduction

La possibilité de détourer un objet avec précision dans une séquence d'images est très importante pour de nombreuses applications de manipulation de séquences vidéo. En effet, de nombreuses activités liées à la post production ou à la réalité augmentée visent à créer des images ou des séquences d'images à partir d'images réelles en ajoutant, en enlevant ou en modifiant des objets dans la scène. Il faut donc être capable de détourer précisément l'objet dans une séquence pour le transformer le manipuler, le colorier ou même le reporter dans une autre séquence vidéo.

A l'heure actuelle, le détourage précis d'un objet dans une image est toujours un problème très complexe qui est soit effectué manuellement, soit avec le concours d'outils semi-automatiques de vision par ordinateur, comme les *intelligent scissors* [10]. Le problème est encore plus complexe dès que le détourage doit être effectué dans une séquence d'images. Il n'est plus envisageable de détourer image par image et il est donc indispensable de disposer de méthodes facilitant la détection de l'objet dans la séquence, à partir d'une détourage initial de l'objet. Nous parlerons alors de suivi. Plusieurs type de méthodes ont été développées pour permettre le suivi d'objets dans une séquence:

- des méthodes purement 2D: ces méthodes supposent que le changement d'apparence de l'objet dans la séquence peut être modélisé par une transformation paramétrique simple (affinité, homographie, modèle d'ordre 2)[9]. Elles sont souvent suivies d'une phase d'adaptation locale pour affiner la précision de l'objet [2, 4, 5]. Cependant ces approches fonctionnent assez mal dès que l'apparence de l'objet change rapidement, comme lorsqu'une nouvelle face de l'objet apparaît.
- des méthodes purement 3D: certains auteurs proposent d'utiliser un dispositif de stéréovision pour obtenir une reconstruction 3D et pouvoir ainsi facilement segmenter la scène [16, 8]. Malgré ce dispositif sophistiqué, la précision des éléments détourés reste assez mauvaise et ne peut être utilisée pour la réalité augmentée. D'autres méthodes, comme [12], cherchent à reconstruire l'objet d'intérêt à partir de plusieurs vues bien choisies, appelées images-clés, dans lesquelles les contours occultants de l'objet ont été précisément délimités. La position de la caméra correspondant à chacune de ces vues est calculée par des techniques devenues maintenant classiques en vision par ordinateur et l'objet est reconstruit comme intersection des cônes d'appuis correspondant à chaque contour détouré. L'objet peut ensuite être déterminé dans chaque image en projetant ce modèle

reconstruit. Cependant, la précision du modèle est très insuffisante; en effet l'imprécision sur le point de vue calculé induit une forte erreur sur le modèle 3D. Ceci conduit à sous estimer largement le volume de l'objet lorsqu'on utilise une longue séquence pour la reconstruction.

La méthode que nous proposons tire parti des ap-

proches 2D et 3D développées jusqu'à présent. Comme

[12], nous utilisons le concept d'images clés ce qui per-

met de prendre en compte les changements d'aspects

de l'objet dans la séquence. Cependant nous opérons

seulement une reconstruction locale de l'objet par une courbe qui ne sera utilisée qu'entre deux images clés. Cette reconstruction est utilisée pour prédire la position de l'image entre deux images clés. Des méthodes de type suivi 2D sous l'hypothèse d'un mouvement paramétrique sont ensuite utilisées pour affiner la détermination de l'objet occultant. Nous obtenons ainsi un système très simple d'utilisation, requérant une interaction très faible de l'utilisateur et permettant d'obtenir un détourage très précis de l'objet dans la séquence. Il est de plus important de souligner que notre système prend explicitement en compte l'erreur d'estimation commise lors du calcul du point de vue pour chaque image de la séquence: nous avons développé une méthode originale permettant le calcul de l'incertitude entachant le calcul du point de vue. Cette erreur est ensuite répercutée au niveau de la reconstruction puis du suivi et permet de contraindre la recherche des contours de l'objet dans les images de la séquence. La section 2 explicite les grandes lignes de notre système et précise les interactions nécessaires avec l'utilisateur. La section 3 montre comment l'erreur sur le point de vue est calculée puis répercutée sur la reconstruction et sur la prédiction de l'objet dans les images. La section 4 précise enfin l'étape d'ajustement permettant d'obtenir un détourage précis de l'objet sur la séquence. Pour terminer, la section 5 montre des exemples de segmentation sur des séquences variées et illustre le potentiel de la méthode pour diverses applications de réalité augmentée.

2 Description du système

2.1 Principe

Notre système, présenté schématiquement figure 1 et illustré sur un exemple réel en figure 2, requiert seulement une intervention modérée de l'utilisateur. Celuici, après avoir déterminé les images où l'objet considéré change d'aspect (comme l'apparition d'une face dans le cas d'un objet polyédrique, voir l'image-clé 2 du schéma), sélectionne ces images comme *image-clés* et détoure l'objet dans ces images (figure 2.a). Comme nous allons le voir, cette tâche, facile à réaliser, va

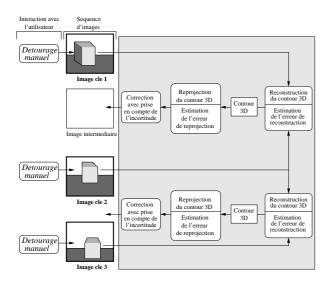


Fig. 1 – Schéma général.

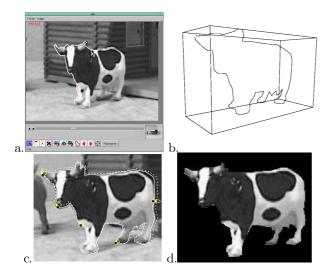


Fig. 2 – Les étapes de notre système. a: détourage dans une image-clé; b: contour 3D reconstruit; c: reprojection du contour3D dans une image intermédiaire et correction; d: objet segmenté.

nous permettre d'obtenir une information fiable sur la structure tridimensionnelle de l'objet.

À partir de ces silhouettes définies par l'utilisateur, on pourrait théoriquement obtenir un modèle tridimensionnel de l'objet (si les images-clés ont été correctement choisies) en calculant l'intersection des cônes s'appuyant sur les contours détourés et dont les sommets sont les positions des caméras [14]. Cependant, si le mouvement de la caméra a été calculé à partir de suivis de points d'intérêt ou de correspondances entre la scène et les images [7, 13], l'imprécision des

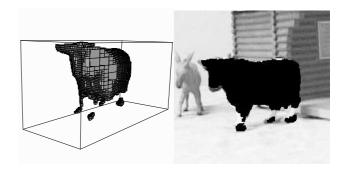


Fig. 3 – Reconstruction obtenue à partir de 5 imagesclés et reprojection dans une image intermédiaire.

points de vue estimés ne permet pas d'obtenir un modèle 3D de l'objet considéré suffisamment précis [12]. Pour s'en persuader, la figure 3.a montre le modèle de la vache reconstruit à partir de 5 images-clés ainsi que la reprojection sur une image intermédiaire (Fig. 3.b): le modèle reconstruit est largement sous estimé et la prédiction est donc assez éloigné de l'objet.

Pour ces raisons, nous n'effectuons qu'un reconstruction locale de l'objet et nous nous limitons à reconstruire un contour 3D correspondant à la silhouette de l'objet pour chaque paire d'images-clés consécutives (figure 2.b). Ces contours sont obtenus par stéréoscopie grâce aux silhouettes définies par l'utilisateur. Il est bien connu que dans le cas d'objets non polyédriques, les deux contours détourés ne correspondent pas à une même courbe 3D. Cependant, nous ne cherchons ici qu'une bonne approximation de la surface de l'objet au voisinage des contours observés: en le reprojetant dans les images intermédiaires correspondant aux deux images-clés, on obtient une prédiction de la silhouette de l'objet dans ces images.

Cette prédiction peut être relativement éloignée de la position attendue s'il s'agit d'un objet non polyédrique, mais surtout à cause du manque de précision sur les points de vue mentionnés plus haut. C'est pourquoi une étape de correction est nécessaire (figure 2.c): nous recherchons alors la transformation affine qui amène le contour reprojeté au niveau de la silhouette de l'objet (figure 2.d). La recherche est réalisée en s'appuyant sur la texture de l'objet dans les images-clés. En outre, une des contributions majeures de cet article est de montrer comment estimer les erreurs de points de vue, problème trop souvent négligé, et comment mettre à profit cette estimation. En effet, à partir de l'estimation des erreurs de points de vue, nous pouvons estimer l'erreur sur la reconstruction des contours 3D, et de là, l'erreur sur la reprojection de ces contours. Cette erreur nous permet le plus souvent d'améliorer les résultats.

2.2 Aspects interactifs

L'intérêt de cette méthode est qu'elle se traduit en un outil précis mais également intuitif. En effet, même si la méthode est basée sur une reconstruction tridimensionnelle des objets, l'utilisateur n'a à considérer que les images de la séquence, sans avoir à fournir d'informations plus complexes que le détourage d'objets 2D. De son point de vue, tout se passe comme si la méthode interpolait les silhouettes entre les images-clés, d'une manière beaucoup plus fiable qu'une interpolation bidimensionnelle puisqu'on prend ici en compte la nature tridimensionnelle des objets.

Le nombre d'images-clés nécessaires dépend du nombre de changements d'aspect de l'objet considéré. En particulier, deux images-clés peuvent suffire quand il n'y a pas de changement d'aspect. La qualité de la segmentation obtenue dépend de la précision des points de vue: s'ils sont trop imprécis, il faut parfois ajouter une image-clé pour améliorer les résultats. Cet ajout peut cependant être réalisé rapidement: il suffit de corriger localement un des contours obtenus dans une image intermédiaire, qui devient alors une nouvelle image-clé.

Enfin, dans notre implémentation, la tâche de détourage dans les images-clés est allégée grâce aux *intelligent scissors*[11] et à des outils plus classiques comme la possibilité de déplacer ou de supprimer un point du contour. Le temps de détourage manuel est généralement compris entre 1 et 2 minutes par image-clé pour des objets relativement complexes.

3 Calcul de l'incertitude sur le point de vue

Nous explicitons dans cette section la méthode nous permettant de déterminer l'incertitude sur les points de vue calculés. D'une manière générale, le point de vue est calculé comme le minimum d'une fonction Φ faisant intervenir des données 3D disponibles sur l'environnement observé et/ou la donnée de points d'intérêt qui peuvent être automatiquement extraits et suivis dans la séquence considérée [7, 13].

Le calcul de l'incertitude sur les points de vue estimés par des méthodes de vision a été peu abordé jusqu'à présent, excepté par [6] qui envisagent le calcul de l'erreur sur la matrice fondamentale. Une approximation de la matrice de covariance est obtenue grâce au théorème des fonctions implicites en supposant que les mesures images nécessaires au calcul sont indépendantes. Cette dernière hypothèse est difficile à assurer et conduit à diminuer arbitrairement la matrice de covariance lorsque le nombre de mesures augmente.

Pour cette raison, nous avons opté pour une méthode

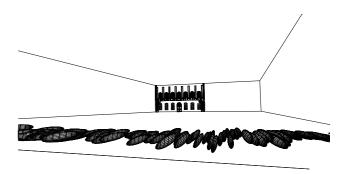


FIG. 4 – Les régions d'indifférence correspondant aux paramètres de translation pour la séquence Stanislas.

plus intuitive, la méthode des régions d'indifférence [1], qui permet d'estimer la fiabilité de la position p estimée de la caméra (p est ici un vecteur à 6 composantes: les 3 angles d'Euler et la translation). L'idée sous tendant les régions d'indifférence est la suivante: étant donné que nous cherchons à minimiser Φ , il semble raisonnable de supposer que des valeurs de Φ presque aussi petites que le minimum atteint Φ^* nous satisferont presque autant que Φ^* . Ce raisonnement conduit à la définition d'une ϵ région d'indifférence:

$$\epsilon_{region} = \{ p \ tel \ que | \Phi(p) - \Phi(p^*) | \le \epsilon \}$$

En développant Φ à l'ordre 1 au voisinage du minimum p^* , on obtient

$$\Phi(p) \approx \Phi(p^*) + \nabla \Phi(p^*)^t \delta p + \frac{1}{2} \delta p^t H(p^*) \delta p$$
(1)

où H^* est le hessien de Φ calculé en $p=p^*$. Comme p^* est le minimum de Φ , $\nabla \Phi(p^*)=0$ et l'équation (1) se réduit donc à:

$$\Phi(p) \approx \Phi(p^*) + \frac{1}{2} \delta p^t H(p^*) \delta p$$

La ϵ région d'indifférence est donc définie par:

$$|\delta p^t H(p^*) \delta p| \le 2\epsilon$$

qui est un ellipsoïde de dimension 6. Un exemple de ces régions d'indifférence, calculé pour la séquence *Stanislas*, est montré en figure 4. Le calcul du point de vue a été réalisé en utilisant le modèle 3D de l'opéra, au fond de la scène. Pour chaque image, nous avons représenté la région d'indifférence correspondant aux paramètres de translation.

La précision sur les points de vue étant maintenant évaluée, il reste à en inférer l'incertitude sur le contour 3D reconstruit, puis l'incertitude sur le contour reprojeté qui servira de prédiction. Si nous disposions de correspondances points à points entre les deux contours détourés, estimer l'imprécision sur le contour reconstruit serait relativement simple [15]. Ici, nous n'avons que des correspondances des contours-clés 1 et 2, et les correspondances de points que nous pouvons établir grâce aux épipolaires dépendent de l'incertitude sur les points de vues calculés. C'est pourquoi nous avons choisi une approche exhaustive pour résoudre ce problème. Nous considérons les poses extrémales c'est à dire les 12 sommets de l'ellipsoïde d'indifférence. No- ${}_{1}^{12}p$ } (resp $\{p_{2}^{1}$ $\binom{12}{2}$) ces poses extrémales tons $\{p_1^1\}$ dans deux vues-clés. Pour tout point de 1, nous considérons toutes les reconstructions possibles de en considérant les 12 positions extrémales correspondant à chacun des points de vue: étant donné p_1^i et p_2 deux positions extrémales, le correspondant de est déterminé en considérant l'intersection de la ligne épipolaire avec 2. On obtient ainsi 144 points reconstruits tenant compte de l'incertitude sur les points de vue. On peut alors considérer que l'enveloppe convexe de ces 144 points est une bonne estimation de l'erreur de reconstruction 3D associée.

L'incertitude sur le contour reprojeté est calculée de la même façon: les 144 points reconstruits sont reprojetés dans l'image intermédiaire en utilisant les 12 positions extrémales de cette vue. L'enveloppe $_i$ de ces points est l'incertitude associée à chaque point du contour prédit. Un exemple de zones d'incertitude dans le cas de la place Stanislas est visible figure 5.

4 Etape d'ajustement

En raison de l'erreur sur les points de vue calculés et également parce que la reconstruction peut être issue de l'observation de contours occultants, la prédiction ainsi obtenue ne correspond en général pas exactement à l'objet recherché. Nous décrivons donc ici l'étape d'ajustement qui va permettre, à partir de la prédiction de l'objet dans les images intermédiaires et de l'incertitude associée, de déterminer le contour effectif de l'objet. En fait, nous pouvons déterminer non seulement le contour prédit mais également sa texture. Ce motif prédit, appelé oti dans la suite, est calculé à partir des images clés en utilisant des transformations locales 2D.

Nous avons retenu une approche hiérarchique souvent utilisée dans le cadre de structures déformables [3]. Nous calculons d'abord une estimation globale de la déformation de la forme entre une vue clé et l'image intermédiaire. Un ajustement local est ensuite réalisé pour ajuster les détails. Les nombreuses expérimentations que nous avons réalisées nous ont montré qu'une transformation affine était bien adaptée pour modéliser la déformation. Nous recherchons donc la meilleure

transformation affine t () = $\begin{cases} 1 & + 2 & + 3 \\ 4 & + 5 & + 6 \end{cases}$ minimisant le critère de corrélation

$$(\) = \sum_{i} \quad (\) \tag{2}$$

οù

$$(\) = \sum_{i=0}^{x} (\ _{oti} \ (\ _{i} +\) -\ (t \ (\ _{i} +\)))^{2}$$

où la courbe prédite est définie comme l'ensemble des points $\{i\}_{1:i:n}$, =(i), est la taille de la fenêtre de corrélation, est la région à l'intérieur de et est l'intensité dans l'image intermédiaire considérée. Une méthode de type Powell est utilisée pour la minimisation de . La convergence de cet algorithme peut cependant être perturbée en particulier lorsque l'objet occultant présente des motifs répétitifs (voir figure 5.a). C'est pourquoi un terme de pénalité est ajoutée à la fonction minimisée de façon à contraindre les points de la courbe à appartenir à la zone d'incertitude i. Le critère optimisé est donc:

$$(\) = \begin{cases} \sum \ (\ _{oti} \ (\ _{i} + \) - \ (t \ \ \ (\ _{i} + \)))^{2} \\ t \ \ (\ _{i}) \ \ i \\ 2 \ (\ \) \end{cases}$$

Cette contrainte nous permet, dans la plupart des cas, d'obtenirla convergence vers le contour recherché (figure 5.b).

5 Résultats

L'efficacité de notre approche est démontrée sur plusieurs séquences. Nous avons fait varier la nature de la trajectoire de la caméra, certaines trajectoires étant moins propices que d'autres à la reconstruction 3D, comme une translation le long de l'axe optique (voir la séquence du chalet). Nous avons également considéré des objets de forme complexe, en particulier non polyédriques. À travers ces résultats, nous montrons également différentes applications de l'outil développé ici.

5.1 Séquence du chalet: occultations en réalité augmentée

Cette séquence est composée de 120 images (figure 6). La caméra se déplace suivant différents mouvements: deux translations suivant l'axe optique et également des rotations. Un modèle 3D partiel du chalet ainsi que la prise en compte du suivi des points

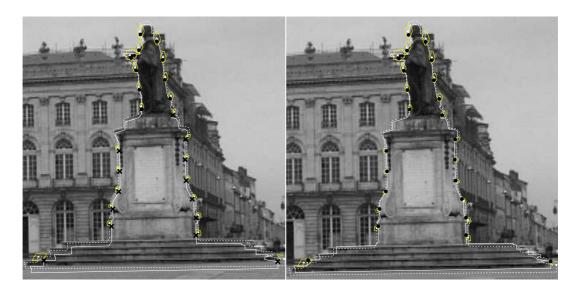


Fig. 5 – Comparaison du suivi obtenu sans (a) et avec (b) la prise en compte des régions d'intérêt. La courbe prédite est en pointillé, la courbe atteinte par le suivi est en traits pleins.

d'intérêt dans la séquence nous a permis d'estimer les points de vue. Nous voulons ajouter dans cette séquence des objets virtuels, en particulier un objet derrière la vache jouet. Pour pouvoir prendre en compte correctement les occultations entre objets réels et virtuels, nous recherchons la silhouette de la vache jouet dans toutes les images de la séquence. La figure 6 montre les images-clés définies. Certaines images-clés sont consécutives: ceci est rendu nécessaire par la complexité du graphe d'aspects de l'objet considéré, en particulier à cause des pattes de la vache jouet. Par exemple, avant l'image 30, seules 3 pattes sont visibles alors que les 4 pattes sont visibles dans l'image 31. Ainsi, la topologie du contour est la même entre les images 0 et 30, de même entre les images 31 et 40, ce qui permet de réaliser la reconstruction 3D des contours, plus précisément la phase d'appariement. Les contours tridimensionnels obtenus sont montrés figure 7. La figure 8 montre ensuite, en pointillés, la reprojection de ces contours dans trois images intermédiaires, ainsi que le contour final obtenu après correction (en trait continu). Enfin, la figure 9 montre la segmentation obtenue ainsi que l'incrustation finale, ce qui permet de valider visuellement les résultats.

5.2 Séquence Stanislas: composition

Cette séquence a été filmée d'une voiture tournant autour de la place Stanislas. Notre but est ici d'extraire la statue pour l'insérer dans une séquence virtuelle. L'estimation des points de vue a été obtenue à l'aide d'un modèle partiel du bâtiment du fond de la scène. Une des faces du piédestal de la statue apparaît image

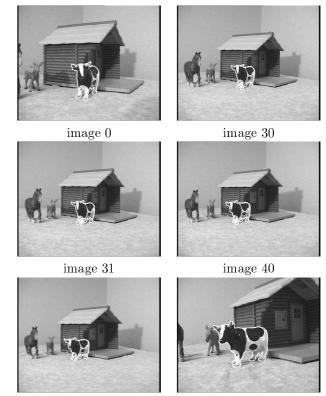


Fig. 6 – Séquence du chalet : les 6 images-clés retenues.

image 120

image 41

72, qui a donc été retenue comme image-clé. La première image-clé est la première image de la séquence

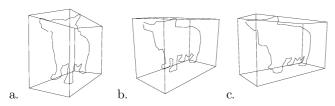
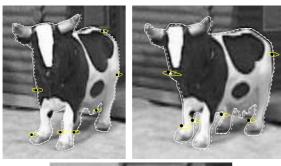


Fig. 7 – Séquence du chalet. Contours 3D reconstruits à partir des images a: 0 et 30; b: 31 et 40 c: 41 et 120.



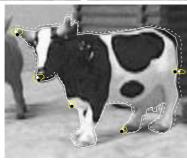


Fig. 8 – Séquence du chalet : reprojection et correction dans des images intermédiaires

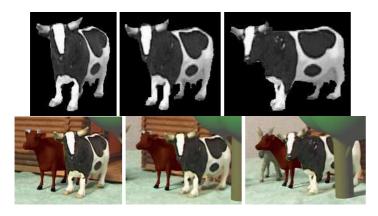


Fig. 9 – Séquence du chalet : résultats (segmentation et incrustation)

où la statue est entièrement visible. La dernière imageclé est la dernière image de la séquence où les marches ne sont pas encore occultées par un piéton (voir figure 10).



Fig. 10 – Séquence Stanislas: les 3 images-clé retenues

La figure 11 montre la prédiction de la silhouette de la statue dans deux images intermédiaires obtenues en reprojetant les contours 3D, ainsi que l'ajustement obtenu. La segmentation de la statue dans ces images et un exemple de scène virtuelle utilisant cette segmentation sont montrées figure 12.

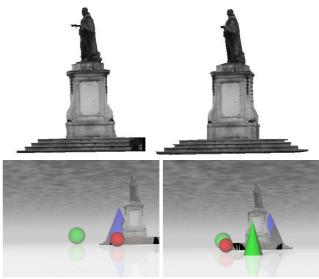
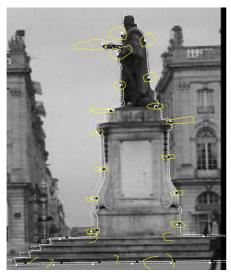


Fig. $12 - S\'{e}quence Stanislas: r\'{e}sultats (segmentation et composition).$

5.3 Séquence des chaînes: réalité diminuée

Cette séquence va nous permettre d'illustrer une autre application de notre outil, la suppression d'un objet dans une séquence vidéo. Un exemple est montré figure 13 pour une séquence dans laquelle nous voulions insérer une course de voitures. Il fallait donc éliminer la chaîne du passage! À partir de deux imagesclés, nous pouvons retrouver rapidement la silhouette de la chaîne sur l'ensemble des images de la séquence (voir figure 13). Après avoir retiré la chaîne de chaque image, comme cet objet est mince, nous pouvoir remplir le trou résultant par interpolation de l'intensité de l'image.



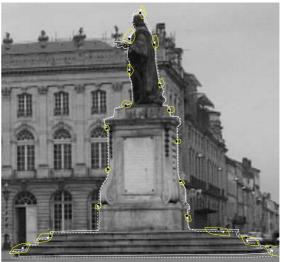


Fig. 11 – Séquence Stanislas: reprojection et correction dans deux images intermédiaires.

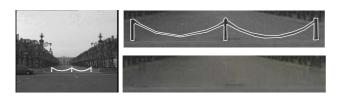


Fig. 13 – Séquence des chaînes: une des deux imagesclé retenues, détourage obtenu dans une image intermédiaire et résultat après interpolation de l'intensité

5.4 Séquence de la voiture: objet mobile et rigide

La méthode que nous avons décrite jusqu'à présent est capable de segmenter un objet qui est fixe par rapport à un référentiel attaché à la scène. En effet, le point de vue est estimé par rapport à un repère lié à la scène et l'objet peut être reconstruit dans ce repère grâce aux images clés sous réserve que l'objet n'ait pas bougé entre les observations.

Nous montrons maintenant que cette méthode peut être utilisée également pour segmenter un objet mobile et rigide: il suffit pour cela de connaître les points de vue dans un référentiel lié à l'objet mobile considéré. Dans la séquence considérée figure 14, où on voit une voiture arriver par la droite de l'image, le mouvement de la caméra par rapport à l'objet mobile (ici la voiture) a été estimé en suivant des primitives 3D sur la voiture (les fenêtres et le coffre). La figure 14 montre les deux images-clés utilisées et le résultat de la segmentation dans une image intermédiaire.



image-clé 12

image-clé 30



Fig. 14 – Séquence de la voiture: images-clés et segmentation dans une image intermédiaire.

6 Conclusion

Nous avons présenté dans ce papier une méthode simple et intuitive permettant la segmentation d'un objet dans une séquence vidéo. Nous montrons ici qu'au prix d'une interaction très faible avec l'utilisateur (le détourage de l'objet dans quelques vues clés), nous pouvons obtenir un détourage très précis de l'objet dans toute la séquence.

Un des points forts de notre méthode est qu'elle utilise une reconstruction 3D locale de la scène en prenant explicitement en compte l'incertitude sur les points de vues calculés ce qui permet de contraindre la recherche de l'objet dans une région d'intérêt.

Sous réserve de bien choisir les images-clés, notre méthode se révèle mieux adaptée et plus précise que la

plupart des méthodes existantes.

Annexe

Les intelligent scissors sont un outil interactif introduit par [11]: l'utilisateur clique tout d'abord à l'aide de la souris sur un point du contour à tracer, point appelé graine. Quand il déplace le pointeur de la souris, l'interface graphique propose un chemin (déterminé par un ensemble de critères, comme la norme du gradient et les niveaux de gris le long de ce chemin) reliant la graine à la position actuelle du pointeur de la souris. Quand le chemin proposé correspond bien à une portion du contour que veut tracer l'utilisateur, celui-ci clique à nouveau pour retenir le chemin proposé, et la position de la souris devient la nouvelle graine. Le processus est itéré jusqu'à ce que l'objet soit entièrement détouré.

Références

- [1] Y. Bard. Nonlinear Parametric Estimation. Academic Press, 1974.
- [2] B. Bascle, P. Bouthemy, R. Deriche, and F. Meyer. Tracking Complex Primitives in an Image Sequence. In Proceedings of the 12th International Conference on Pattern Recognition, Jerusalem (Israel), 1994.
- [3] B. Bascle and R. Deriche. Stereo Matching Reconstruction and Refinement of 3D curves Using Deformable Contours. In *Proceedings of 4th International Conference on Computer Vision*, Berlin (Germany), pages 421–430, 1993.
- [4] M.-O. Berger. How to Track Efficiently Piecewise Curved Contours with a View to Reconstructing 3D Objects. In Proceedings of the 12th International Conference on Pattern Recognition, Jerusalem (Israel), volume 1, pages 32–36, 1994.
- [5] M. Covell and J. Darrell. Dynamic occluding contours. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Colorado (USA), June 1999.
- [6] Csurka, C. Zeller, Z.Y. Zhang, and O.D. Faugeras. Characterizing the Uncertainty of the Fundamental matrix. Computer Vision and Image Understanding, 68(1):18–36, May 1997.
- [7] A.W. Fitzgibbon and A. Zisserman. Automatic Camera Recovery for Closed or Open Images Sequences. In *Proceedings of 5th European Confe*rence on Computer Vision, University of Freiburg (Germany), pages 311–326, June 1998.
- [8] G. Gordon, T. Darrell, M. Harville, and J. Woodfill. Background estimation and removal based on range and color. In *Proceedings of IEEE Confe-*

- rence on Computer Vision and Pattern Recognition, Colorado (USA), June 1999.
- [9] F. Meyer and P. Bouthemy. Region Based Tracking in an image sequence. In Proceedings of 2nd European Conference on Computer Vision, Santa Margherita Ligure (Italy), volume 588 of Lecture Notes in Computer Science, pages 476–483, 1992.
- [10] E. Mortensen and W. Barrett. Intelligent Scissors for Image Composition. In *Computer Graphics* (*Proceedings Siggraph*), pages 191–198, 1995.
- [11] E. Mortensen and W. Barrett. Interactive Segmentation with Intelligent Scissors. *Graphical Models and Image Processing*, 60(5):349–384, September 1998.
- [12] K. C. Ong, H. C. Teh, and T. S. Tan. Resolving Occlusion in Image Sequence Made Easy. *The* Visual Computer, 14:153–165, 1998.
- [13] G. Simon and M.-O. Berger. Des méthodes efficaces pour l'incrustation d'objets virtuels dans des séquences d'images. *Traitement du Signal*, 16(1):31–46, 1999.
- [14] R. Szeliski. Rapid Octree Construction from Image Sequences. CVGIP: Image Understanding, 58(1):23–32, July 1993.
- [15] J.I. Thomas, A. Hanson, and J. Oliensis. Understanding Noise: The Critical Role of Motion Error in Scene Reconstruction. In Proceedings of 4th International Conference on Computer Vision, Berlin (Germany), 1993.
- [16] M. Wloka and B. Anderson. Resolving Occlusions in Augmented Reality. In Symposium on Interactive 3D Graphics Proceedings, (New York), pages 5–12, August 1995.