Zones quasi-plates spatio-temporelles et segmentation morphologique de s equences vid eo

Jonathan Weber, Sébastien Lefèvre, Pierre Gancarski

Zones quasi-plates spatio-temporelles et segmentation morphologique de séquences vidéo

Jonathan Weber[1,*] Sébastien Lefèvre[2] Pierre Gançarski[1]
[1] LSIIT - Université de Strasbourg
[2] VALORIA - Université de Bretagne Sud
[*] Pôle API - Bd Sébastien Brant - BP 10413
67412 Illkirch CEDEX FRANCE
j.weber@unistra.fr

8 avril 2011

Résumé

La qualité d'une segmentation s'apprécie généralement au regard de l'usage qui en est fait. Afin de s'adapter aux besoins très variés pour lesquels elle est employée, la segmentation peut être guidée par l'utilisateur, au lieu d'être complètement automatique. La morphologie mathématique a fournit plusieurs méthodes de segmentation guidée par l'utilisateur, reposant le plus souvent sur la Ligne de Partage des Eaux. Néanmoins, Soille a récemment suggéré une nouvelle approche consistant à assembler des pièces de puzzle obtenues en produisant les zones quasi-plates (ZQP) d'une image. Dans cet article, nous étudions plus profondément ce schéma de segmentation guidée par l'utilisateur dans le contexte des séquences vidéo. Nous introduisons ainsi le concept de ZQP spatio-temporelles, et proposons plusieurs méthodes pour extraire de telles zones d'une séquence vidéo.

Mots-clés

Zones quasi-plates, segmentation vidéo, personnalisation de segmentation

Abstract

Segmentation quality often depends on the problem under consideration. In order to face the various needs for which it is involved, segmentation may be driven by the user, instead of being fully automatic. Mathematical Morphology has provided several user-driven segmentation approaches, mostly relying on the watershed transform. Nevertheless, Soille has recently suggested another solution consisting in gathering puzzle pieces computed as Quasi-Flat Zones (QFZ) of an image. In this paper, we study more deeply this user-driven segmentation scheme in the context of video data. Thus we introduce the concept of Spatio-Temporal QFZ and propose several methods for extracting such zones from a video sequence.

Keywords

Quasi-flat zones, video segmentation, segmentation personalization

1 Introduction

Après l'augmentation massive des données textuelles, et plus récemment des images, disponibles dans des bases de données et sur le Web, nous observons aujourd'hui une augmentation des données vidéo. De nombreux traitements ou utilisations de telles données nécessitent souvent une segmentation préalable afin d'obtenir des objets d'intérêt sur lesquels effectuer ces traitements. La segmentation d'une vidéo n'est pas unique et dépend des besoins de l'utilisateur. Il est donc nécessaire de pouvoir disposer d'une méthode de segmentation permettant la personnalisation du résultat.

Les méthodes de segmentation vidéo issues de la Morphologie Mathématique sont réparties en deux catégories : les méthodes automatiques [2, 4] qui ne nécessitent aucune interaction avec l'utilisateur (hors réglage éventuel de paramètres) et les méthodes interactives [5, 6, 7, 9] où l'utilisateur fournit des marqueurs pour chaque objet d'intérêt afin de guider la segmentation. Les résultats obtenus par les méthodes automatiques produisent souvent une sur-segmentation et ne sont généralement pas adaptés aux besoins de l'utilisateur. Les méthodes interactives nécessitent une implication plus forte de l'utilisateur, mais fournissent un résultat personnalisé. Cependant, dans le cadre de la segmentation vidéo, les marqueurs initiaux ne fournissent généralement pas à la première tentative la segmentation désirée par l'utilisateur. Il faut donc les corriger, puis relancer le processus de segmentation dans son intégralité. Ces approches peuvent donc s'avérer très coûteuses en temps de calcul. Une autre solution pour résoudre le problème de la personnalisation des segmentations est de fournir une sur-segmentation que

l'utilisateur va réduire en fusionnant les régions afin d'obtenir la segmentation qu'il désire. Une méthode dans ce cas peut-être de partitionner l'image en zones plates [10], c'est à dire en composantes connexes dont les pixels ont la même valeur. Une frontière entre deux objets étant généralement située entre des pixels de valeur différente, les frontières des objets sont inclues dans les frontières des zones plates, en assemblant les zones plates on peut donc obtenir n'importe quel objet. Cependant, les zones plates produisent une sur-segmentation extrême. Afin de réduire cette sur-segmentation tout en conservant les propriétés intéressantes des zones plates, les zones quasi-plates (ZQP) ont été proposées. Les ZQP sont basées sur un critère de construction moins restrictif qui conduit à la production de zones plus grandes, tout en assurant un cout calculatoire faible et un ensemble de frontières susceptible de contenir la majorité des frontières des objets d'intérêt. Soille [12] précise par ailleurs que les ZQP ne sont pas réellement des méthodes de segmentation mais plutôt des méthodes qui décomposent une image en pièces de puzzle. Les ZQP sont alors une étape de pré-traitement dans un processus de segmentation d'image basée sur la fusion de pièces de puzzle. La fusion de ces différentes pièces peut-être guidée par l'utilisateur, ce qui permet d'offrir la caractéristique intéressante de personnalisation des résultats. Cependant, dans le cas des séquences vidéo, il n'existe encore aucune définition des ZQP. Le but de ce papier est de proposer une telle définition et d'étudier son application dans le cadre de la segmentation guidée par l'utilisateur.

Dans cet article, nous allons rappeler les concepts inhérents aux ZQP tels que formulés dans le cadre de la connexité des prédicats logiques introduite par Soille [11, 14]. Puis, nous étendrons cette définition aux séquences vidéo. Nous étudierons ensuite son application à la personnalisation de segmentation vidéo et montrerons son intérêt par des premiers résultats. Nous conclurons sur les perspectives offertes par cette approche.

2 Segmentation d'images par zones quasi-plates

2.1 Connexité des prédicats logiques

Les ZQP sont basées sur la notion de chemin α-connexe. Un chemin est α-connexe si tous les chemins de deux pixels le composant sont Lipschitz-continus, ou autrement dit :

Un chemin $\mathcal{P}$, selon un voisinage N, composé de n pixels (p_0,p_1,…,p_{n-1}) est un chemin α-connexe (α-$\mathcal{P}$) si et seulement si :

$$\forall i \in [0,n-2], p_i \in N(p_{i+1}) \text{ et } |f(p_i) - f(p_{i+1})| \leq \alpha \quad (1)$$

Cette notion permet de définir les ZQP les plus simples, c'est-à-dire les zones α-connexes [8] que nous noterons α-$\mathcal{Z}$ et définies par :

$$\alpha\text{-}\mathcal{Z}(p) = \{p\} \cup \{Q \mid \forall q \in Q, \alpha\text{-}\mathcal{P}(p,q) \neq \emptyset\} \quad (2)$$

où α-$\mathcal{P}(p,q)$ est l'ensemble des chemins α-connexes entre p et q. L'α-$\mathcal{Z}$ d'un pixel p est donc l'ensemble des pixels auquel il est relié par un chemin α-connexe. On note que les zones plates sont un cas particulier de l'α-$\mathcal{Z}$ avec $\alpha = 0$. Les α-$\mathcal{Z}$ ont la propriété hiérarchique suivante qui nous sera utile par la suite :

$$\forall \alpha' \leq \alpha, \alpha'\text{-}\mathcal{Z}(p) \subseteq \alpha\text{-}\mathcal{Z}(p) \quad (3)$$

Notons que la segmentation d'une image en α-$\mathcal{Z}$ peut produire une sous-segmentation : en effet, si on augmente trop la valeur de α, il peut apparaitre une *réaction en chaine*, qui peut mener selon les images et la valeur d'α à obtenir une seule ZQP pour toute l'image. Afin de contrer ce phénomène, de nouvelles définitions de ZQP basées sur l'α-$\mathcal{Z}$ ont été développées. Dans un but d'unification de ces différentes définitions, Soille et Grazzini [11, 14] ont proposé un cadre théorique, la *connexité des prédicats logiques*. Nous rappelons qu'un prédicat logique P renvoie $vrai$ quand l'argument satisfait le prédicat, $faux$ sinon. Soille et Grazzini définissent donc un nouveau type de ZQP qu'on note (P_1,…,P_n)-$\mathcal{Z}$ qui permet de produire des ZQP vérifiant les n prédicats logiques P_1,…,P_n. Ces prédicats peuvent être de diverses natures, citons par exemple le prédicat de variation globale ω (qui vérifie si la différence entre les valeurs minimale et maximale des pixels au sein d'une ZQP est inférieure à un seuil) ou encore l'indice de connexité β (qui est vérifié si le ratio entre le nombre de chemins α-connexes de deux pixels et le nombre de chemins de deux pixels au sein d'une ZQP est supérieur à un seuil). La (P_1,…,P_n)-$\mathcal{Z}$ consiste donc, pour chaque pixel p, à chercher la plus grande α-$\mathcal{Z}$ satisfaisant tous les prédicats. Grâce à la propriété (3), nous savons que si $\alpha' < \alpha$ alors l'α'-$\mathcal{Z}(p)$ est plus petite ou égale à l'α-$\mathcal{Z}(p)$. Si les prédicats ne sont pas vérifiés pour une valeur de α, le constat précédent permet de décrémenter α afin de vérifier s'ils sont vérifiés pour une valeur inférieure jusqu'à trouver la valeur maximale de α pour laquelle tous les prédicats sont vérifiés. Les auteurs donnent la formulation mathématique suivante pour la connexité des prédicats logiques :

$$\begin{aligned}(P_1,\ldots,P_n)\text{-}\mathcal{Z}(p) = \bigvee \{ & \alpha'\text{-}\mathcal{Z}(p) \mid \\ & \forall k \in \{1,\ldots,n\}, \forall \alpha\text{"} \leq \alpha', \forall q \in \alpha'\text{-}\mathcal{Z}(p), \\ & P_k(\alpha'\text{-}\mathcal{Z}(p)) = vrai \text{ et } P_k(\alpha\text{"}\text{-}\mathcal{Z}(q)) = vrai\}\end{aligned} \quad (4)$$

Ce cadre théorique est adapté aux méthodes garantissant un résultat unique ou, autrement dit, assurant la

propriété d'unicité : en effet, on cherche l'α'-$\mathcal{Z}$ la plus grande satisfaisant tous les prédicats logiques. Il n'est donc pas possible d'intégrer les méthodes ne produisant pas une segmentation en ZQP unique. Plus qu'un cadre permettant d'unifier les définitions existantes, la $(P_1,\ldots,P_n)$-$\mathcal{Z}$ permet également de développer de nouvelles définitions de ZQP. Actuellement trois prédicats sont utilisés dans les ZQP, la variation locale (α), la variation globale (ω) et l'indice de connexité (β). Dans le cadre défini par Soille et Grazzini, nous pouvons intégrer des prédicats portant sur d'autres caractéristiques (périmètre, aire, etc.) mais également sur des descripteurs plus complexes (variation de texture, gradient, etc.) sous réserve que ces prédicats respectent la définition (4).

Enfin, ce cadre théorique a été défini pour les images en niveaux de gris, et ne peut donc être appliqué tel quel aux images multibandes telles que les images couleur, alors que ces dernières représentent la majorité des images disponibles actuellement. Cependant, des indications pour l'utilisation des zones quasi-plates dans des espaces multibandes ont été données par Soille [12]. Ce dernier fixe la contrainte d'un paramètre α défini comme un vecteur ayant la même valeur dans chaque bande. Ainsi on peut aisément hiérachiser les α et donc disposer d'un ordre total (par exemple, le décrément de α = (3,3,3) est (2,2,2)). La contrainte d'avoir la même valeur dans chaque bande peut être contournée en pré-traitant individuellement chaque bande pour en adapter les valeurs. Le prédicat de variation globale est quant à lui traité comme celui de variation locale : il est satisfait seulement s'il est vérifié marginalement pour chacune des bandes.

Dans la suite de cet article, nous ferons l'hypothèse que les ZQP seront des ZQP couleur produites par la $(P_1,\ldots,P_n)$-$\mathcal{Z}$ en ne tenant compte que des prédicats de variation locale et globale.

2.2 Filtrage

Les ZQP souffrent d'un problème inhérent aux régions de transition. Ces dernières sont les régions situées entre deux objets, et où se manifeste un phénomène d'escalier sur les valeurs des pixels frontaliers. Cet escalier est introduit par la discrétisation de l'image et l'interpolation des valeurs qu'elle entraine. Ce phénomène provoque une sur-segmentation à proximité de cette frontière, qui se retrouve alors composée de ZQP de très petite taille. Des solutions ont été proposées pour résoudre ce problème. Soille et Grazzini [14] définissent les régions de transition comme les ZQP ne contenant que des pixels de transition. Un pixel de transition est ici assimilé à un pixel qui n'est pas un extrémum local. Les auteurs proposent alors de supprimer toutes les ZQP correspondant à des régions de transition, puis à accroitre les ZQP restantes en utilisant un algorithme de croissance de régions [1]. Après

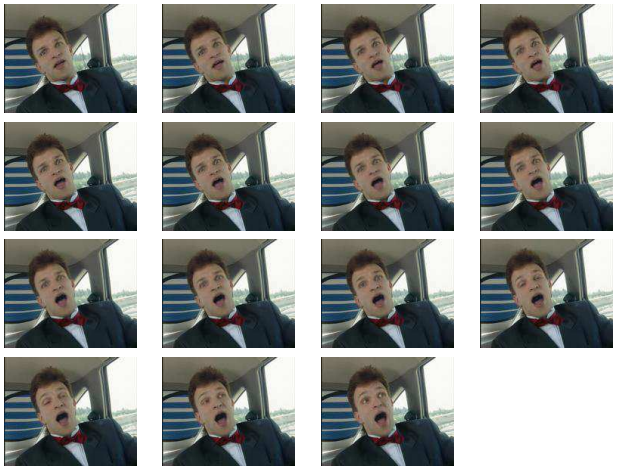

FIGURE 1 – Trames de l'extrait de *carphone*.

la suppression de ces régions, le nombre de zones plates est réduit de façon importante.

La solution proposée par Soille et Grazzini ne dépend d'aucun paramètre et repose sur une définition précise de ce qu'est une région de transition. Mais on observe que subsistent dans l'image de nombreuses régions ne contenant que quelques pixels. Ces régions ne correspondent pas à la définition des régions de transitions, et provoquent pourtant une forte sur-segmentation. Nous pensons donc que, même si la méthode proposée dans [14] réduit la sur-segmentation en supprimant les régions de transition, elle est encore insuffisante puisque ces régions de transition ne sont pas les seules responsables de la sur-segmentation.

D'autres auteurs ont proposé des méthodes de filtrage des ZQP basées sur l'utilisation d'un seuil d'aire minimale des ZQP. Une méthode, proposée par Angulo et Serra [3], fusionne toutes les ZQP ayant une aire inférieure à un seuil avec les ZQP voisines les plus similaires. Avec ce procédé, aucune région de transition n'est présente dans la segmentation finale. Zanoguera [15] supprime les ZQP dont l'aire est inférieure à un seuil (incluant ainsi les régions de transition) avant d'appliquer une ligne de partage des eaux pour accroitre les ZQP restantes dans les zones où se trouvaient les ZQP supprimés. Soille [13] propose une méthode de filtrage basée sur une augmentation itérative de l'aire minimale : à chaque itération, les ZQP dont l'aire est supérieure ou égale à l'aire minimale sont utilisées dans un algorithme de croissance de régions, et on assigne ensuite une valeur unique (par exemple la valeur moyenne) à chacune des régions obtenues. La simplification de l'image obtenue est segmentée en ZQP à l'itération suivante, le processus étant répété jusqu'à ce que les ZQP filtrées ne changent plus. En nous inspirant de ces méthodes, nous proposons une nouvelle méthode de filtrage qui se démarque de l'existant par son efficacité algorith-

mique. Notre méthode s'appuie sur l'algorithme Seeded Region Growing (SRG) [1] mais, au lieu de l'appliquer sur des pixels, nous l'appliquons sur les ZQP. Pour ce faire, nous utilisons un seuil d'aire minimale à l'instar des méthodes précédentes. Nous produisons une segmentation en ZQP que nous transformons en graphe d'adjacence de régions. Nous considérons toutes les ZQP dont l'aire est supérieure ou égale à ce seuil comme des graines pour le SRG que nous appliquons au graphe d'adjacence de régions (le principe est illustré sur un exemple artificiel figure 2). Notons que plus le seuil d'aire est élevé, plus la sur-segmentation est réduite, mais plus le risque d'obtenir une sous-segmentation de certains objets d'intérêt est élevé. Par contre, la croissance de régions étant effectuée sur les ZQP et non sur les pixels, notre méthode de filtrage est peu couteuse en temps de calcul.

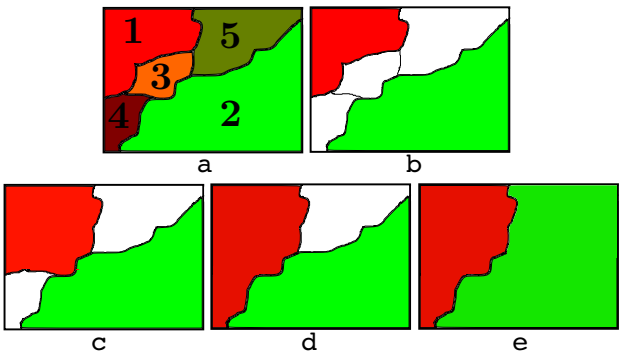

FIGURE 2 – Notre méthode de filtrage de ZQP : a) ZQP originales, b) Suppression des ZQP dont l'aire est inférieure au seuil, c) 1ère itération du SRG : la ZQP 1 croît en incorporant les pixels de la ZQP 3, d) 2ème itération du SRG : la ZQP 1 croît en incorporant les pixels de la ZQP 4, e) Dernière itération du SRG : la ZQP 2 croît en incorporant les pixels de la ZQP 5

3 Extension des ZQP aux séquences video

3.1 Limites de l'extension dite $3D$

L'extension la plus directe des ZQP aux séquences vidéo est de considérer une séquence vidéo comme un cube spatio-temporel. On peut ainsi réutiliser les définitions existantes, puisque la seule différence porte sur le voisinage à considérer (ici un voisinage spatio-temporel et non plus seulement spatial).

En appliquant la $(P_1,\ldots,P_n)$-$\mathcal{Z}$ en $3D$, on obtient une sur-segmentation spatiale beaucoup plus élevée qu'en $2D$. Sur l'extrait de *carphone* (figure 1) pour $\alpha = \omega = 20$ (à des fins de comparaison, nous conserverons ces valeurs dans la suite de l'article), on obtient en moyenne 4441 $\mathcal{Z}$ par trame en $2D$ contre 6779 $\mathcal{Z}$ en $3D$ (soit pour la segmentation complète des 15 trames : 55040 $\mathcal{Z}$). Ce phénomène s'explique par le fait qu'en traitant les volumes en 3D, le voisinage comporte plus de pixels, et donc une α-$\mathcal{Z}$ comportera plus de pixels (cf. *réaction en chaîne* discutée précédemment), ce qui augmentera naturellement le risque de violer un ou plusieurs des prédicats considérés. Ainsi, la plus grande α-$\mathcal{Z}$ satisfaisant tous les prédicats est souvent celle produite avec une valeur de α faible, ce qui conduit à obtenir de très petites régions ne comportant que quelques pixels. Nous constatons donc que les définitions de ZQP utilisant des contraintes (ou prédicats) supplémentaires semblent peu utilisables en $3D$.

Si traiter les séquences vidéo comme des volumes $3D$ semble naturel, les définitions actuelles de ZQP sont totalement inadaptées à un traitement $3D$ des séquences vidéo. D'autres extensions doivent être envisagées.

3.2 Traitement séquentiel des dimensions spatiales et temporelle

L'approche $3D$ n'étant pas adaptée aux séquences vidéo, nous nous intéressons ici à une approche consistant à traiter successivement (et non plus conjointement) les dimensions spatiales et temporelle. La figure 3 résume cette approche. Nous allons dans un premier temps étudier l'approche spatial vers temporel ($2D+t$), puis nous étudierons l'approche temporel puis spatial ($t+2D$).

L'approche spatial vers temporel consiste à analyser, dans un premier temps, chaque trame indépendamment des autres pour en extraire les ZQP. Nous considérons ensuite les ZQP comme les noeuds d'un graphe que nous valuons d'un attribut représentatif des ZQP (par exemple la valeur moyenne des pixels de chaque ZQP). Nous relions ensuite temporellement les ZQP d'une trame aux ZQP des trames adjacentes (précédente et suivante) dont les pixels se superposent spatialement. Chaque arête est alors valuée de la différence entre les valeurs des deux noeuds qu'elle relie. Les nouvelles ZQP sont donc les plus grandes composantes connexes de noeuds reliés par des arêtes dont la valeur est inférieure ou égale à α et qui ne violent aucun prédicat. On constate que la $(P_1,\ldots,P_n)$-$\mathcal{Z}$ utilisant le prédicat de variation globale produit nettement moins de régions en $2D+t$ (23926 $\mathcal{Z}$) qu'en $3D$ (55040 $\mathcal{Z}$), réduisant ainsi d'autant le phénomène de sur-segmentation.

L'obtention de résultats meilleurs en $2D+t$ qu'en $3D$ s'explique par le traitement différent des deux types de dimensions. En $3D$, les dimensions spatiale et temporelle ne sont pas différenciées mais à l'inverse traitées de la même façon, ce qui accroit la sur-segmentation dans le cadre de la $(P_1,\ldots,P_n)$-$\mathcal{Z}$. A l'opposé, en $2D+t$, le premier traitement sur la seule dimension

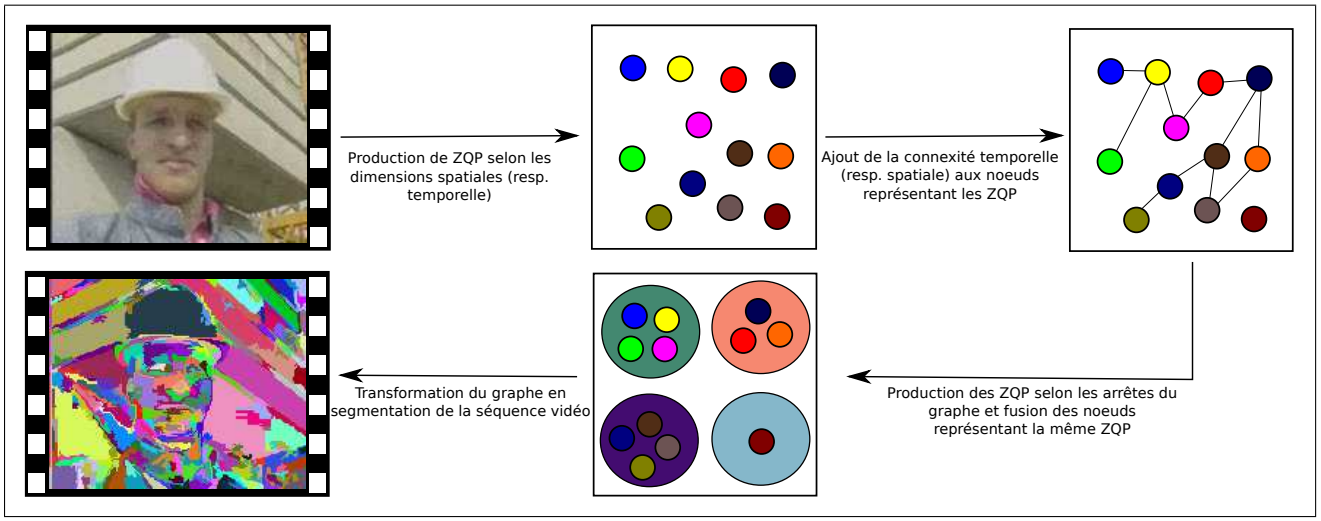

FIGURE 3 – Segmentation d'une séquence vidéo en zones quasi-plates spatio-temporelles par traitement séparé des dimensions spatiale et temporelle.

spatiale maximise la taille des ZQP puisqu'il limite le voisinage : rappelons que plus le voisinage est important, plus le risque d'avoir un pixel qui met en échec un prédicat est grand, ce qui réduira la valeur α de l'α-$\mathcal{Z}$ maximale et donc la taille de la $(P_1,\ldots,P_n)$-$\mathcal{Z}$. Le traitement spatial permet donc d'agrandir l'aire des ZQP, réduisant ainsi la sur-segmentation spatiale. Par contre, le traitement temporel va pour sa part entrainer une sur-segmentation temporelle. Cette sur-segmentation est liée aux contraintes des prédicats : les régions étant spatialement plus étendues, elles sont moins homogènes, et peuvent donc avoir des valeurs moyennes suffisamment différentes pour mettre en échec un prédicat lors du calcul temporel des $(P_1,\ldots,P_n)$-$\mathcal{Z}$.

L'approche temporel puis spatial consiste à produire, pour chaque coordonnée spatiale prise indépendamment, les ZQP selon la dimension temporelle. On obtient donc, après le traitement temporel, une sur-segmentation spatiale extrême puisque, pour chaque trame, chaque pixel appartient à une région différente. A l'instar de ce qui a été fait en $2D+t$, nous transformons les ZQP temporelles obtenues en noeuds d'un graphe. Ces noeuds sont valués d'un attribut, par exemple la valeur moyenne des pixels qu'ils représentent. Nous relions ensuite les ZQP spatialement voisines par des arêtes, et valuons ces dernières de la différence entre les deux ZQP qu'elles relient. À partir de ce graphe sont enfin produites les ZQP finales qui sont les plus grands ensembles connexes de noeuds reliés par des arêtes dont la valeur est inférieure ou égale à α et qui respectent tous les prédicats. On constate que la $(P_1,\ldots,P_n)$-$\mathcal{Z}$ produit ici moins de régions (16830 $\mathcal{Z}$) qu'avec l'approche $2D+t$, notamment grâce à une sur-segmentation temporelle plus réduite.

La $(P_1,\ldots,P_n)$-$\mathcal{Z}$ avec le prédicat de variation globale met en lumière un phénomène intéressant : de par l'ordre dans lequel elles traitent les différentes dimensions, les approches $2D+t$ et $t+2D$ induisent des sur-segmentations différentes ; l'approche $2D+t$ provoque une sur-segmentation spatiale réduite mais une forte sur-segmentation temporelle ; à l'inverse, l'approche $t+2D$ produit une sur-segmentation spatiale forte mais une sur-segmentation temporelle réduite. Les deux approches produisent néanmoins des résultats de meilleure qualité que l'approche $3D$. Le choix de l'une ou l'autre des approches dépendra vraisemblablement de la séquence vidéo à traiter : l'approche $2D+t$ semble plus adaptée à des séquences vidéo de haute résolution spatiale mais d'une durée relativement courte, tandis que l'approche $t+2D$ semble plus adaptée à des séquences vidéo de faible résolution mais d'une durée plus longue.

Quelle que soit la dimension considérée (spatiale ou temporelle), nous avons systématiquement utilisé ici la $(P_1,\ldots,P_n)$-$\mathcal{Z}$. Cette dernière garantissant l'unicité du résultat, et cette propriété étant conservée par composition, les approches $2D+t$ et $t+2D$ garantissent donc l'unicité du résultat.

3.3 Filtrage des ZQP vidéo

La méthode basée sur un seuil d'aire minimale présentée dans la section 2.2 peut être étendue en ne considérant plus une aire minimale mais un volume minimal. Cependant, le choix d'un volume minimal, même s'il semble efficace dans le contexte d'images réellement tri-dimensionnelles, n'est pas adapté aux séquences vidéo qui sont de nature spatio-temporelle et non purement spatiale. Ainsi, prenons le cas d'une

ZQP n'étant composée que de quelques pixels au sein de chaque trame (soit d'une aire faible) mais présente sur de nombreuses trames successives : celle-ci sera conservée de par son volume supérieur au seuil de volume minimal, alors qu'elle ne représente sans doute pas un objet mais une partie d'un objet. Face à ce constat, nous proposons d'utiliser plutôt un seuil d'aire moyenne minimale, l'aire moyenne aire* étant calculée comme :

$$\text{aire}^*(\alpha\text{-}\mathcal{Z}) = \frac{\text{aire}(\alpha\text{-}\mathcal{Z})}{\text{long}(\alpha\text{-}\mathcal{Z})} \quad (5)$$

où long représente la durée d'une ZQP, ou autrement dit le nombre de trames successives où celle-ci est présente dans la séquence vidéo.
A l'instar des filtrages effectués sur les images fixes, la sur-segmentation est ici fortement réduite : en considérant une aire minimale de 10 pixels, nous obtenons 686 $\mathcal{Z}$ en $3D$, 845 $\mathcal{Z}$ en $2D+t$ et 378 $\mathcal{Z}$ en t+2D. De plus, nous n'observons peu voire pas de sous-segmentation : les ZQP obtenues par cette méthode semblent ainsi utilisables dans un contexte de segmentation.
Le filtrage par seuil d'aire minimale est très efficace pour réduire la sur-segmentation des ZQP dans les séquences vidéo. On obtient une réduction très importante de la sur-segmentation tout en conservant la qualité des ZQP. En combinant les définitions de ZQP adaptées aux séquences vidéo et le filtrage par aire, nous obtenons une méthode efficace de pré-segmentation. Cette pré-segmentation pourra être finalement exploitée par des méthodes de fusion de ZQP dans le but d'obtenir la segmentation désirée par un utilisateur donné.

4 Segmentation vidéo guidée par l'utilisateur

4.1 Méthode

Un problème classique en segmentation vidéo est la sur-segmentation : par exemple, la méthode de Ligne de Partage des Eaux (LPE) prédictive [4], appliquée à l'extrait de *carphone* (figure 1), produit un résultat de façon non-supervisée qui comporte environ 2000 régions. Ce problème est également présent dans la segmentation par zones quasi-plates : en effet, la segmentation de la même séquence par la $(P_1,\ldots,P_n)$-$\mathcal{Z}$ $t+2D$, avec $\alpha = \omega = 20$ et une aire minimale de 10 pixels, produit une sur-segmentation (378 $\mathcal{Z}$ pour les 15 trames) ainsi qu'un résultat également non personnalisé. L'introduction d'une interaction avec l'utilisateur va permettre de personnaliser la segmentation et de réduire la sur-segmentation. La segmentation guidée par l'utilisateur est un principe connu en Morphologie Mathématique. Dans le contexte des ZQP, elle peut prendre la forme d'un assemblage guidé des pièces de puzzle associées aux différentes ZQP. Cette approche a pour but de fournir à l'utilisateur une méthode intuitive et générique de segmentations de vidéos.

Méthode	**Paramètres**	**Précision**		
		(a)	**(b)**	**(c)**
$(P_1,\ldots,P_n)$-CC $2D+t$	$\alpha=\omega=10$	.837	.982	.981
$(P_1,\ldots,P_n)$-CC $2D+t$	$\alpha=\omega=20$	.841	*.987*	.989
$(P_1,\ldots,P_n)$-CC $2D+t$	$\alpha=\omega=30$	*.851*	.981	.989
$(P_1,\ldots,P_n)$-CC $2D+t$	$\alpha=\omega=40$	*.882*	*.987*	.989
$(P_1,\ldots,P_n)$-CC $t+2D$	$\alpha=\omega=10$	**.899**	.968	.970
$(P_1,\ldots,P_n)$-CC $t+2D$	$\alpha=\omega=20$	.828	.979	.984
$(P_1,\ldots,P_n)$-CC $t+2D$	$\alpha=\omega=30$	.814	**.988**	**.993**
$(P_1,\ldots,P_n)$-CC $t+2D$	$\alpha=\omega=40$	.837	*.986*	.988
LPE guidée marqueurs		.851	.985	.990
Seeded Region Growing		.802	.806	.824

TABLE 1 – Comparaison de la précision pixel en considérant différents marqueurs (a) quelques points sur la trame médiane, b) marqueurs épais sur la trame médiane, c) marqueurs épais sur trois trames.

Nous nous inspirons de ce principe d'interaction en proposant une méthode de segmentation en ZQP guidée par des marqueurs, dont le principe est le suivant : dans un premier temps, les ZQP initiales sont produites de façon automatique (non-supervisée) ; l'utilisateur fournit ensuite des marqueurs (par exemple à l'aide d'une interface lui permettant de les dessiner sur l'image), ce qui lui permet ainsi de personnaliser la segmentation en indiquant les objets qu'il désire ; les ZQP couvertes par les marqueurs sont alors considérées comme les graines d'un algorithme de Seeded Region Growing [1], qui va fusionner les différentes ZQP selon leur distance en termes de couleur, ce qui peut être assimilé à un critère α dans le contexte des ZQP. L'utilisateur ne visualisant pas les ZQP initiales, il est possible que plusieurs marqueurs chevauchent une même ZQP : dans ce cas, la ZQP et les marqueurs sont incompatibles et nous devons procéder à une correction. Nous supposons que l'utilisateur a correctement fourni les marqueurs, et considérons donc que la ZQP est incorrecte et doit être segmentée à nouveau : pour cela, on applique l'algorithme Seeded Region Growing sur la ZQP en utilisant les marqueurs comme graines ; on corrige alors la ZQP, améliorant ainsi la précision de la sur-segmentation initiale.

4.2 Expérimentations

Afin de valider la pertinence de cette approche, nous avons mené des expérimentations sur l'extrait de la séquence *carphone*. Nous avons donc comparé les segmentations guidées exploitant les $(P_1,\ldots,P_n)$-$\mathcal{Z}$ $2D+t$ et $t+2D$ à deux méthodes classiques dont nous avons réutilisé les principes : d'une part la LPE guidée par des marqueurs et d'autre part la méthode de Seeded Region Growing. Ainsi nous pouvons évaluer notre

FIGURE 4 – Marqueurs utilisés pour la séquence *carphone* : (a) quelques points sur la trame 7, (b) marqueurs épais sur la trame 7, (c) à (e) marqueurs épais sur les trames 3, 7 et 10.

contribution vis-à-vis de l'emploi direct de l'une ou l'autre de ces approches classiques. Pour cela, nous avons utilisé trois configurations différentes de marqueurs (présentées en figure 4), les trois jeux de marqueurs ont été réalisés indépendamment. Les résultats sont donnés dans le tableau 1, où la précision correspond au ratio de pixels bien segmentés (affectés à la bonne région). Nous avons utilisé différents jeux de paramètres (α,ω) pour les $(P_1,\ldots,P_n)$-$\mathcal{Z}$ afin d'étudier l'impact de ces paramètres sur les résultats. Notons que, quelque soit les marqueurs utilisés, notre méthode donne toujours de meilleurs résultats qu'un SRG. Cela montre que notre méthode en appliquant un SRG sur les ZQP est plus performante qu'une application directe du SRG sur les pixels. Concernant la comparaison de la LPE guidée par des marqueurs, notre méthode produit de meilleurs résultats sur chaque ensemble de marqueurs, mais pas avec tous les jeux de paramètres. En outre, il n'y a pas de combinaison de paramètres unique donnant de meilleurs résultats avec tous les ensembles de marqueurs. Cependant, si l'on exclut la première série de marqueurs (quelques points sur la trame médiane) qui sont nettement insuffisants, nous observons que $(P_1,\ldots,P_n)$-$\mathcal{Z}$ $t+2D$ et $\alpha=\omega=30$ produit les meilleurs résultats avec les deux autres ensembles de marqueurs. A l'instar des deux autres méthodes, la segmentation interactive par ZQP est très sensible aux marqueurs. Ceci est illustré par la figure 5, qui montre les résultats obtenus par la $(P_1,\ldots,P_n)$-$\mathcal{Z}$ $2D+t$ avec $\alpha=\omega=40$ et $aire^*=10$ sur les différents ensembles de marqueurs.

Nous avons également comparé notre proposition à une méthode récente de segmentation morphologique interactive, la LPE à propagation de marqueurs [5]. Pour cette comparaison, nous avons choisi de définir des marqueurs sur la première trame. Les paramètres considérés pour la LPE à propagation de marqueurs sont la liaison de marqueurs et la propagation de

FIGURE 5 – Résultats de la $(P_1,\ldots,P_n)$-$\mathcal{Z}$ $2D+t$ basées marqueurs avec les paramètres $\alpha=\omega=40$ et $aire^*=10$ selon différents jeux de marqueurs, (en haut) quelques points sur la trame 8, (milieu) marqueurs épais sur la trame 8, (en bas) marqueurs épais sur les trames 4, 8 et 11.

Méthode	**Paramètres**	**Précision**
$(P_1,\ldots,P_n)$-CC $2D+t$	$\alpha=\omega=10$	*0.985*
$(P_1,\ldots,P_n)$-CC $2D+t$	$\alpha=\omega=20$	*0.984*
$(P_1,\ldots,P_n)$-CC $2D+t$	$\alpha=\omega=30$	0.978
$(P_1,\ldots,P_n)$-CC $2D+t$	$\alpha=\omega=40$	0.979
$(P_1,\ldots,P_n)$-CC $t+2D$	$\alpha=\omega=10$	0.964
$(P_1,\ldots,P_n)$-CC $t+2D$	$\alpha=\omega=20$	0.966
$(P_1,\ldots,P_n)$-CC $t+2D$	$\alpha=\omega=30$	*0.983*
$(P_1,\ldots,P_n)$-CC $t+2D$	$\alpha=\omega=40$	**0.988**
LPE à propagation de marqueurs		0.983

TABLE 2 – Comparaison de la précision pixel de la $(P_1,\ldots,P_n)$-$\mathcal{Z}$ selon différents paramètres et de la LPE par propagation de marqueurs.

mouvement basée régions. Nous n'avons pas permis l'édition de marqueurs puisque l'objectif était de dresser une comparaison dans des conditions équivalentes (ici le temps nécessaire à l'utilisateur). Les résultats sont présentés dans le tableau 2 : on y observe que la $(P_1,\ldots,P_n)$-$\mathcal{Z}$ est, pour certains paramètres, plus précise que la LPE à propagation de marqueurs dans ces conditions. Cela signifie que même sans utiliser l'information de mouvement, la $(P_1,\ldots,P_n)$-$\mathcal{Z}$ guidée par marqueurs permet d'obtenir des résultats comparables à une méthode utilisant cette information.

5 Conclusion

Dans cet article, nous avons proposé une extension des ZQP aux séquences vidéo, ainsi qu'une méthode pour assembler ces ZQP afin d'obtenir une segmentation personnalisée par l'utilisateur. Le traitement séparé des dimensions spatiale et temporelle améliore la segmentation par rapport à un traitement tridimensionnel de la séquence. La méthode proposée

pour assembler les différentes ZQP en fonction des besoins de l'utilisateur est intuitive. Elle fournit en outre de résultats intéressants, par comparaison avec d'autres méthodes de la littérature.

Nos prochains travaux vont porter sur l'amélioration des marqueurs : en effet, la vidéo est pour l'instant marquée avant le traitement ; or, il est intéressant de pouvoir corriger *a posteriori* les marqueurs, à l'instar de ce qui est fait dans [5]. Ceci est d'autant plus pertinent qu'au travers de la mise à jour éventuelle de certaines ZQP, la sur-segmentation peut être améliorée par les marqueurs. Notons que le cout calculatoire principal de notre approche provient de la création des ZQP, et que la fusion basée sur les marqueurs est quant à elle rapide puisqu'effectuée sur le graphe d'adjacence des ZQP. Le temps de calcul suite aux interactions avec l'utilisateur est donc limité, contrairement à d'autres méthodes de segmentation interactives qui nécessitent de relancer le processus de segmentation dans son ensemble (cf. LPE et SRG). Nous envisageons également l'application des ZQP spatio-temporelles à de nouveaux espaces de réprésentation des données, en calculant par exemple les ZQP non plus sur les valeurs initiales des pixels mais directement sur le flot optique. Nous projetons en outre de travailler à l'amélioration du processus de fusion des ZQP en utilisant des descripteurs plus robustes et plus pertinents des ZQP (au lieu d'une simple couleur moyenne). Enfin, notre méthode étant basée sur un processus de réduction de graphe, nous souhaitons étudier comment celui-ci peut permettre l'apprentissage de segmentations. Autrement dit, notre objectif serait de pratiquer un apprentissage à partir de quelques séquences vidéo marquées par l'utilisateur, afin de permettre la segmentation de vidéo non-marquées mais simplement sur-segmentées avec des ZQP.

Remerciements

Ce travail a été soutenu par Ready Business System et l'Association Nationale de la Recherche et de la Technologie (ANRT). Nous remercions particulièrement Christian Dhinaut de RBS pour sa contribution.

Références

[1] R. Adams et L. Bischof : Seeded region growing. *IEEE Transanction on Pattern Analysis and Machine Intelligence*, 16(6):641–647, 1994.

[2] V. Agnus : *Segmentation spatio-temporelle de séquences d'images par des opérateurs de morphologie mathématique.* Thèse de doctorat, Université Louis Pasteur, Strasbourg, 2001.

[3] J. Angulo et J. Serra : Color segmentation by ordered mergings. *In Proceedings of the IEEE International Conference on Image Processing*, pages 125–128, 2003.

[4] S.-Y. Chien, Y.-W. Huang et L.-G. Chen : Predictive watershed : a fast watershed algorithm for video segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(5): 453–461, Mai 2003.

[5] F.C. Flores et R.A Lotufo : Watershed from propagated markers : An interactive method to morphological object segmentation in image sequences. *Image and Vision Computing*, 28(11): 1491–1514, 2010.

[6] C. Gu et M-C. Lee : Semiautomatic segmentation and tracking of semantic video objects. *IEEE Transactions on Circuits and Systems for Video Technology*, 8(5):572–584, 1998.

[7] B. Marcotegui, F. Zanoguera, P. Correia, R. Rosa, R. Mech et M. Wollborn : A video object generation tool allowing friendly user interaction. *In IEEE International Conference on Image Processing*, volume 2, pages 391–395, 1999.

[8] M. Nagao, T. Matsuyama et Y. Ikeda : Region extraction and shape analysis in aerial photographs. *Computer Graphics and Image Processing*, 10(3):195–223, 1979.

[9] J.-F. Rivest, S. Beucher et J. Delhomme : Marker-controlled segmentation : an application to electrical borehole imaging. *Journal of Electronic Imaging*, 1(2):136–142, 1992.

[10] J. Serra et P. Salembier : Connected operators and pyramids. *In Proceedings of SPIE, Non-Linear Algebra and Morphological Image Processing,*, volume 2030, pages 65–76, 1993.

[11] P. Soille : On genuine connectivity relations based on logical predicates. *In Proceedings of the 14th International Conference on Image Analysis and Processing*, pages 487–492, Washington, DC, USA, 2007. IEEE Computer Society.

[12] P. Soille : Constrained connectivity for hierarchical image partitioning and simplification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(7):1132–1145, Juillet 2008.

[13] P. Soille : Constrained connectivity for the processing of very-high-resolution satellite images. *International Journal of Remote Sensing*, 31(22): 5879–5893, 2010.

[14] P. Soille et J. Grazzini : Constrained connectivity and transition regions. *In Proceedings of the 9th International Symposium on Mathematical Morphology and Its Application to Signal and Image Processing*, pages 59–69, Berlin, Heidelberg, 2009. Springer-Verlag.

[15] F. Zanoguera : *Segmentation interactive d'images fixes et de séquences vidéo basée sur des hierarchies de partitions.* Thèse de doctorat, Ecole des Mines de Paris, 2001.