Résumé vidéo par regroupement de seams

MARC DECOMBAS[1,2], FREDERIC DUFAUX[1], BEATRICE PESQUET-POPESCU[1]

[1] Dépt. TSI – Télécom Paris Tech, 37-39 rue Dareau, 75014, Paris, France
[2] Lab. MultiMedia – Thales Communications & Security, 4 Avenue des Louvresses, Gennevilliers, France
marc.decombas@thalesgroup.com, dufaux@telecom-paristech.fr, pesquet@telecom-paristech.fr

Résumé - Des millions de caméras de surveillance sont déployées dans les rues, les aéroports, etc. Des outils de résumé vidéo sont devenus indispensables pour traiter rapidement et à faible coût cette immense quantité d'informations. Nous proposons ici une approche de résumé vidéo par *seam carving* qui permet d'ajouter des contraintes sur les *seams* une fois ceux-ci calculés, ce qui laisse plus de flexibilité et permet de faire un résumé qui s'adapte mieux au contenu. Notre approche permet de faire rentrer dans la scène un objet alors que le précédent n'en est pas encore sorti et cela en diminuant très fortement les anachronismes temporels et les déformations géométriques des objets d'intérêt. Nous obtenons un très bon taux de réduction temporelle tout en préservant les objets d'intérêt.

Summary: Millions of videos cameras have been deployed in streets, airports, etc. Tools for video summary become essential for quickly managing this huge amount of data at a low cost. We propose in this article a new approach of video summary based on seam carving that add constraints on the seams after having computing them, that let more flexibility and allow to have a summary better adapted to the content. Our approach allows an object to enter the scene while the previous one is still present. This is achieved while strongly decreasing temporal anachronisms and geometric deformations on salient objects. A good temporal rate of reduction is reached while preserving salient objects.

1 Introduction

Depuis quelques années, des millions de caméras de surveillance ont été déployées dans les rues, les aéroports, les centres commerciaux. En 2007, le nombre de caméras de surveillance a atteint les 30 millions aux Etats Unis, produisant plus de 4 milliards de séquences vidéo chaque semaine [1]. Le développement d'algorithmes qui identifient automatiquement ce qui est important et qui créent un résumé vidéo devient indispensable pour les opérateurs. Cet article se concentre sur les algorithmes qui réalisent automatiquement des résumés vidéo pour des applications de vidéo surveillance. Le résumé vidéo est la façon d'obtenir une vidéo plus courte où l'ensemble des objets importants y est représenté et où les zones sans intérêt sont supprimées. Ce type de vidéo peut également être utilisé pour avoir un résumé vidéo d'un film personnel ou encore pour des applications législatives [2].

Trois grandes approches sont possibles pour réaliser un résumé vidéo. La première approche est l'accélération de la vidéo, où des trames sont supprimées sur un intervalle fixe ou adaptatif [3][4][5]. La principale limitation de cette approche est que l'on ne peut supprimer que des trames entières. Le taux de réduction y est donc relativement faible, le taux de réduction étant défini comme le ratio de la longueur de la vidéo d'origine par rapport à la longueur de la vidéo traitée. La deuxième approche est le résumé vidéo par extraction de trames clés et représentation de ses trames de manière simultanée sous forme d'un ensemble de vignettes [2]. Le problème de cette approche est que l'aspect temporel est perdu et par la même occasion le contexte. La dernière approche est le montage vidéo, où des segments spatio-temporels des objets d'intérêt sont extraits et combinés pour réaliser un puzzle temporel. Dans [6] et [7], cette approche est appelée synopsis vidéo et n'autorise que des translations temporelles. Elle peut vite devenir complexe et peut créer une inversion des événements lors de la combinaison des objets dans le résumé vidéo.

Pour réaliser ce puzzle, de travaux récents utilisent une méthode appelée *seam carving*. Elle a été proposée en premier par Avidan dans [8] pour réaliser du redimensionnent d'image adapté au contenu. L'idée est de supprimer des *seams* qui sont des chemins de connexité 8 définis verticalement ou horizontalement. Ces *seams* sont calculés à partir d'une carte d'intérêt qui met en évidence les éléments importants de l'image. Dans [9], Rubinstein étend cette approche à des applications vidéo et propose une nouvelle méthode pour calculer les *seams*, qui prend en compte l'influence de la suppression de ceux-ci. Une application simple du *seam carving* sans contrainte pour faire du résumé vidéo crée des anachronismes temporels, des déformations des objets et un résumé n'ayant pas la même longueur sur toutes les lignes. Dans [10] et [11], il propose de calculer les *seams* un à un et la contrainte spatio-temporelle est appliquée lors du calcul de ces *seams*, ce qui peut mener à des résumés vidéo non optimaux. Dans [13], nous utilisons le *seam carving* sur le plan (x,y) avec un regroupement permettant de faire de la compression vidéo bas débit. Nous proposons dans cet article de faire du résumé vidéo en utilisant du *seam carving* par regroupement spatio-temporel contraint. Afin de faire la réduction temporelle, le *seam carving* est appliqué sur le plan (x,t) avec une réduction de t.

Notre approche calcule d'abord l'ensemble des *seams* et analyse ensuite leurs évolutions dans l'espace et le temps. Nous proposons (1) une méthode regroupement spatio-temporelle efficace qui permet (2) de définir le taux de réduction temporel en fonction du contenu, (3) de supprimer les groupes de *seams* isolés, (4) d'identifier des groupes de *seams* spatio-temporels assez grands et (5) d'approximer par segments constants le nombre de *seams* par groupes, tout en ayant le nombre de *seams* total constant. Ceci permet d'éviter les déformations géométriques, les anachronismes et d'avoir un résumé de la même longueur sur toutes les lignes. Nos contraintes laissent plus de flexibilité aux *seams* qui s'adaptent mieux au contenu et permettent d'obtenir un bon taux de réduction temporelle tout en préservant les objets d'intérêt.

2 *Seam carving* temporel par groupes

En considérant une vidéo comme un cube (*N*,*M*,*T*), le *seam carving* peut être appliqué dans le plan (*x*,*t*) avec une réduction de t permettant de faire du résumé vidéo. Cependant, sans aucune contrainte, des anachronismes visuels, des déformations géométriques et des problèmes de longueur de vidéo peuvent apparaitre. Les méthodes précédentes ont proposé d'ajouter des contraintes sur les *seams* dès leurs calculs, ce qui ne permet pas de s'adapter correctement aux situations. Nous proposons donc de calculer en premier l'ensemble des *seams* dans le plan (*x*,*t*) pour ensuite effectuer notre regroupement spatio-temporel contraint.

2.1 L'approche générale

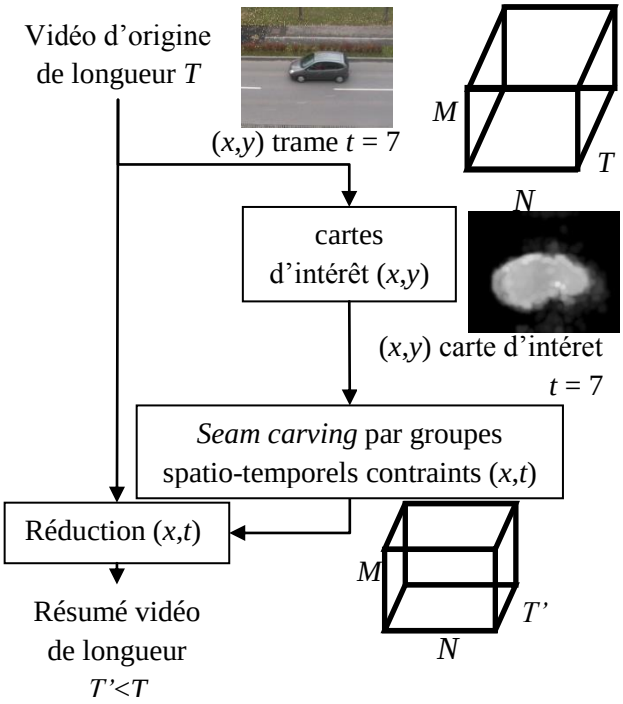


Figure 1 : Approche générale de résumé vidéo par groupes de *seams* temporels contraints

La Figure 1 représente notre approche de résumé vidéo par groupes de *seams* temporels contraints. A partir d'une vidéo d'origine, des cartes d'intérêt sont créées à l'aide du modèle ST-RARE [12] calculé dans le plan (*x*,*y*). Ce modèle identifie les objets d'intérêt en cherchant la rareté sur un ensemble de cartes et en n'utilisant que les plus pertinentes pour les combiner et obtenir une unique carte d'intérêt. A l'origine, ce modèle utilise des composantes statiques (L,a,b) et dynamiques (direction et sens du mouvement) afin d'identifier les zones d'intérêt lorsqu'il y a du mouvement ou non. Or lorsqu'il n'y a pas de mouvement, le modèle identifie des zones d'intérêt statiques, ce qui empêche la suppression de trame. Nous ne prenons donc que l'aspect dynamique du modèle pour effectuer du résumé vidéo. Une fois l'ensemble des cartes d'intérêt calculé, le *seam carving* par groupes spatio-temporels contraints de la Figure 2 est appliqué dans le plan (*x*,*t*) et permet d'obtenir une liste de *seams* supprimables dans chaque trame (*x*,*t*). Ces *seams* permettant de réduire *t* sont supprimés dans la vidéo d'origine afin d'obtenir une vidéo résumée de longueur $t' < t$.

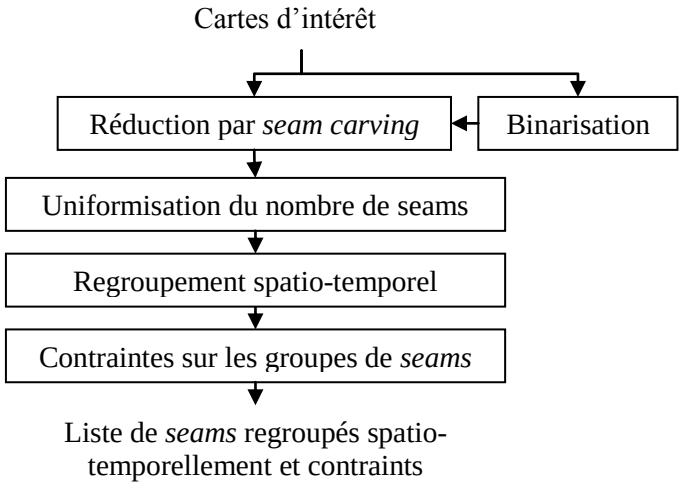


Figure 2 : *Seam carving* par groupes temporels contraints

2.2 Taux de réduction temporelle

Le *seam carving* est un processus itératif permettant de passer d'une résolution à une autre en fonction d'un critère d'arrêt. L'objectif ici est de réduire autant que possible l'aspect temporel tout en préservant les objets d'intérêt. Pour arrêter le processus de réduction, nous utilisons une binarisation de la carte d'intérêt comme critère d'arrêt. Tant qu'un *seam* ne traverse pas la carte binaire, la trame (*x*,*t*) passe à (*x*,*t*-1). Le *seam carving* est appliqué indépendamment sur chaque trame (*x*,*t*) et s'arrête en fonction de l'activité. Puisque l'objectif est de maintenir toutes les zones d'intérêt, et comme certaines trames présentent plus d'activité que d'autres, le taux de réduction temporelle est défini comme le nombre minimal de *seams* supprimables sur l'ensemble des trames (*x*,*t*). Cette étape est le bloc d'« uniformisation du nombre de *seams* » de la Figure 2.

2.3 Les groupes de *seams*

Si le *seam carving* est appliqué sans contrainte spatio-temporelle, des artefacts peuvent apparaitre, comme il est possible de le voir sur l'image (c) des Figure 5, Figure 6 et Figure 7. Notre approche résout ce

problème en créant des groupes de *seams*. Cela permet d'identifier les *outliers* et de contrôler la variation du nombre de *seams* dans chaque groupe. Pour effectuer le regroupement spatio-temporel, les *seams* sont d'abord regroupés spatialement à l'aide d'une *Distance_Spatiale* et d'un *Seuil_Spatial*. La *Distance_Spatiale* est définie comme :

$$D(Seam_t, Seam_{t+1}) = \max_{i=1..N}(Seam_t(i), Seam_{t+1}(i)),$$

avec N la longueur du *seam*. Le *Seuil_Spatial* représente la distance maximale entre deux *seams* que l'on peut avoir dans le même groupe et il a été défini expérimentalement à 7 pixels. Ensuite, les groupes doivent être liés temporellement entre eux. Pour cela on utilise la différence symétrique entre les aires des groupes de la trame *t* avec les groupes de la trame *t+1* et on les regroupe en fonction du minimum de cette différence. Ces étapes sont dans le bloc « regroupement spatio-temporel » de la Figure 2. Ensuite, certaines contraintes sont rajoutées sur les groupes dans le bloc « contraintes sur les groupes de *seams* » de la Figure 2. Les groupes non présents au sein d'assez de trames sont supprimés et leurs *seams* sont réalloués dans les autres groupes. Le nombre de *seams* par groupes varie linéairement en fonction du temps ($Nb\_Seam_{Gpe\_x}=f_{Gpe\_x}(t)$). Or, pour éviter les anachronismes, il faut que cette fonction f_{Gpe_x} soit constante par morceaux. La longueur de ces morceaux étant définie par la taille des objets d'intérêt et le temps qu'ils restent dans la scène. Pour chaque fonction $f_{Gpe\_x,}$ on applique d'abord un filtre médian en fonction du temps, puis on fait une détection de rupture qui permet de segmenter la fonction en morceaux. La valeur médiane est associée à chaque morceau. On obtient ainsi, pour chaque groupe, un ensemble de morceaux constants. Il faut ensuite vérifier que le nombre de *seams* total est toujours constant. Si ce n'est pas le cas, les groupes récupèrent ou suppriment des *seams*. Ceci est illustré dans la Figure 3.

3 Quelques résultats visuels et conclusion

Pour évaluer notre approche, nous avons choisi une séquence de vidéo surveillance avec une caméra fixe et des objets d'intérêt (voitures, vélos) passant de gauche à droite ou de droite à gauche avec beaucoup de trames sans objet. Ce type de séquence est très représentatif de cas réels de vidéos surveillance.

On peut observer sur la Figure 4 les *seams* supprimables (en rouge) après la permutation à y = {139, 154, 244,253}. Sur la première et la deuxième image, on a la trajectoire de 4 véhicules sur la partie supérieure de la route allant de gauche à droite, et sur la troisième et la quatrième image, on a la trajectoire d'un véhicule allant de droite à gauche.

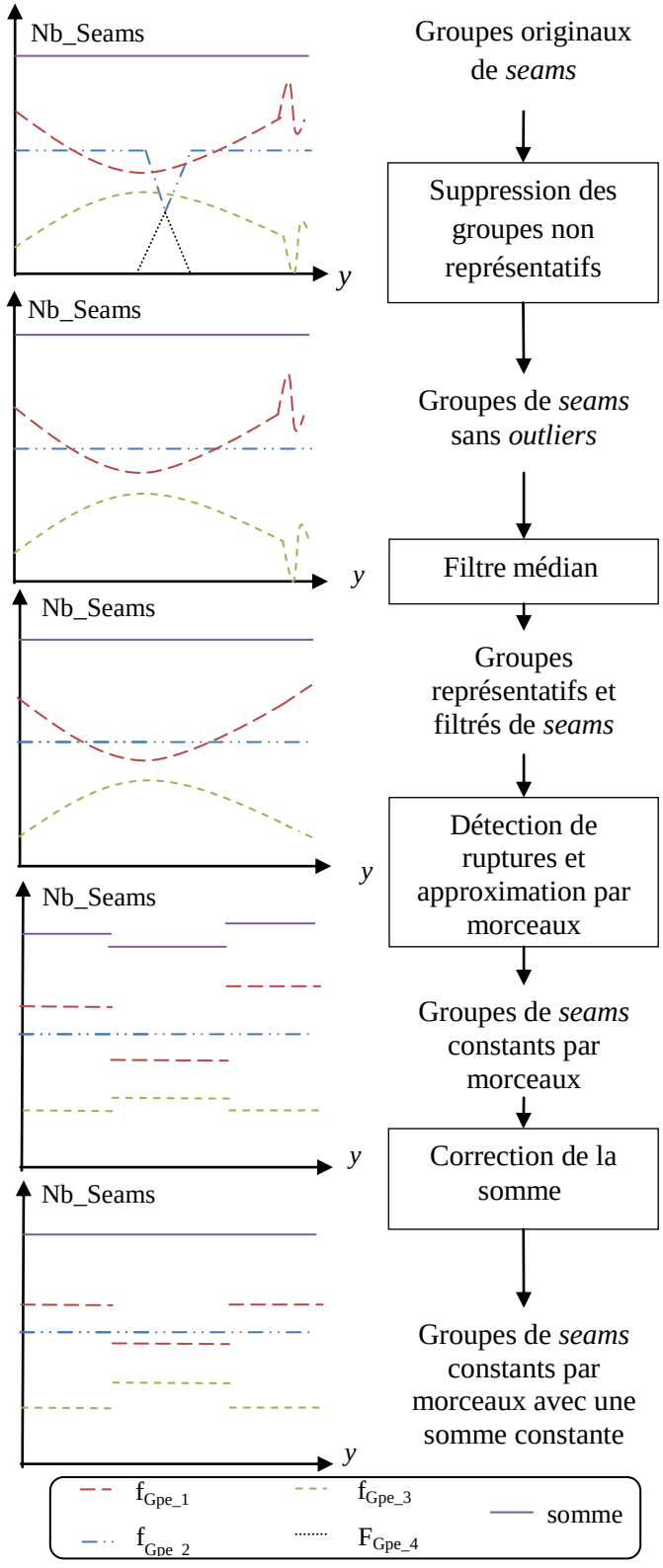
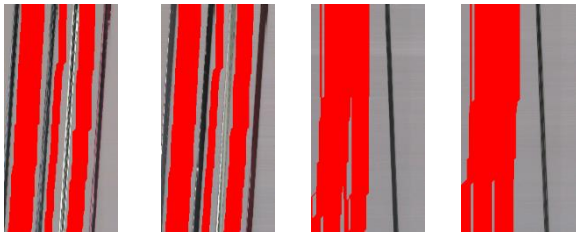


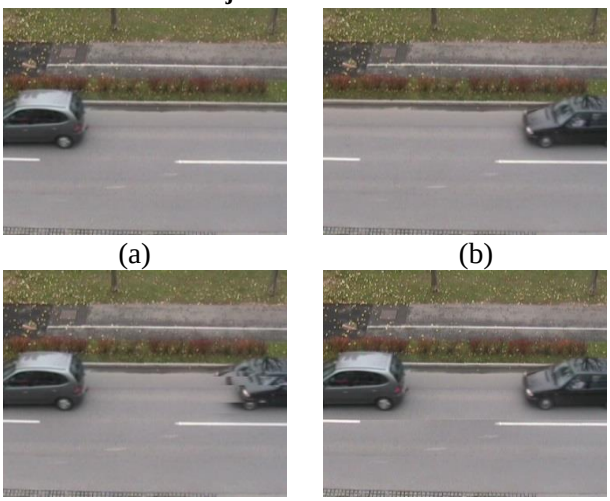
Figure 3 : Contrainte sur les groupes de seams. Des groupes originaux aux groupes constants par morceaux

Grace à notre approche, la quantité de *seams* entre les véhicules reste constante, comme on peut le voir entre y = 139 et y = 154 ou entre y = 244 et y = 253. Les *seams* peuvent quand même s'adapter en fonction du nombre d'objets et de leur trajectoire, comme on peut le voir sur l'ensemble des trames permutées.



(a) (b) (c) (d)

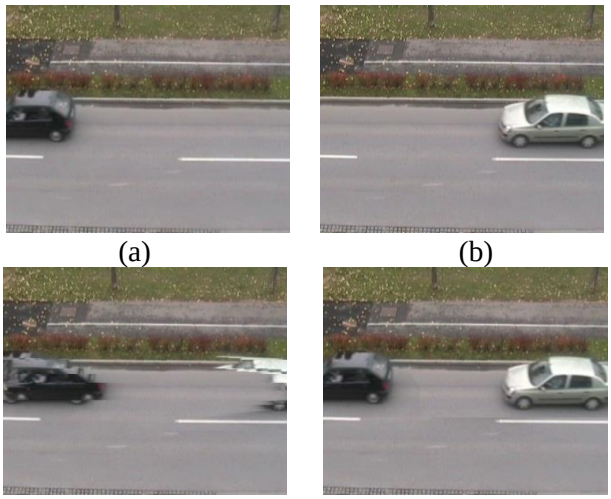
Figure 4 : Visualisation du plan (*x*,*t*) pour y= {139, 154, 244, 253} avec en rouge les *seams* supprimables. Le gris est la route et en noir, blanc et rouge, on retrouve les trajectoires des différents véhicules.



(a) (b)

(c) (d)

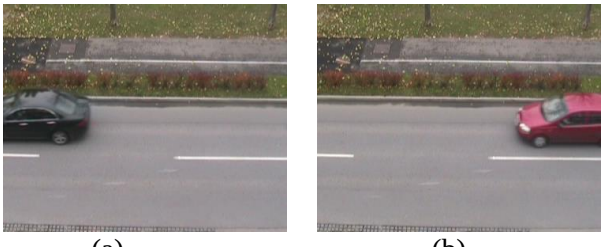
Figure 5 : Résumé vidéo A : (a) trame d'origine à *t* = 17, (b) trame d'origine à *t* = 58, (c) trame après *seam carving* sans contrainte à *t* = 17, (d) trame après *seam carving* par groupes temporels contraints à *t* = 17.



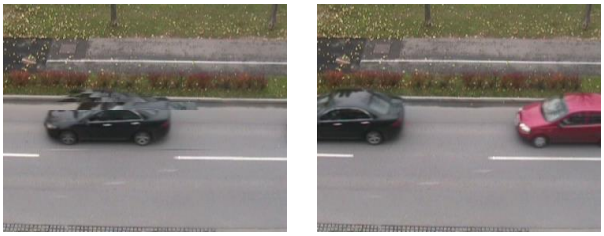
(a) (b)

(c) (d)

Figure 6 : Résumé vidéo B: (a) trame d'origine à *t* =77, (b) trame d'origine à *t* = 94, (c) trame après *seam carving* sans contrainte à *t*= 38, (d) trame après *seam carving* par groupes temporels contraints à *t* = 17



(a) (b)



(c) (d)

Figure 7 : Résumé vidéo C: (a) trame d'origine à *t* = 29, (b) trame d'origine à *t* = 85, (c) trame après *seam carving* sans contrainte à *t* = 20, (d) trame après *seam carving* par groupes temporels contraints à *t* = 20

On peut observer sur la Figure 5, Figure 6 et Figure 7 les résultats de notre approche. Sur les deux premières trames (a) et (b), on voit deux véhicules dans les trames d'origine avec leur temps de passage respectif. Une approche par *seam carving* sans contrainte donne la troisième image (c). On voit de manière générale que les véhicules ont bien été rapprochés dans le temps mais des artefacts géométriques sont apparus. Ceci est dû au fait que le nombre de *seams* supprimés avant, entre et après les véhicules n'est pas constant sur toutes les lignes. Des morceaux du véhicule sont plus «en avance » que d'autres. Dans notre approche, le nombre de *seams* étant constant par morceaux, toutes les lignes des véhicules sont avancées de la même manière, comme on peut le voir sur la dernière image (d). Ceci limite très fortement la quantité d'artéfacts géométriques.

4 Bibliographie

[1] J. Vlahos, "Welcome to the planopticon," *Popular Mechanics*, vol 185, no. 1, pp. 64–69, Jan.2008

[2] J. Oh, Q. Wen, J. Lee, and S. Hwang, "Video abstraction," in *VideoData Mangement and Information Retrieval*, S. Deb, Ed. Hershey, PA: Idea Group, Inc./IRM Press, pp. 321–346, chap. 3, 2004.

[3] M.Yeung, and B.-L.Yeo, "Video visualization for compact presentation and fast browsing of pictorial content," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 5, pp. 771–785, Oct. 1997.

[4] J. Nam, and A. Tewfik, "Video abstract of video," in *Proc. IEEE Workshop on MultiMedia Signal Processing (MMSP'99)*, pp. 117–122, 1999

[5] N. Petrovic, N. Jojic, and T. Huang, "Adaptive video fast forward," *Multimedia Tools Appl.*, vol. 26, no. 3, pp. 327–344, Aug. 2005.

[6] Y. Pritch, A. Rav-Acha, and S. Peleg, "Non-chronological video synopsis and indexing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 11, pp. 1971 – 1984, Nov. 2008.

[7] Y. Pritch, S. Ratovitch, A. Hendel, and S. Peleg, "Clustered Synopsis of Surveillance Video", *6th IEEE Int. Conf. on Advanced Video and Signal Based Surveillance*, Genoa, Italy, Sept. 2009

[8] S. Avidan, and A. Shamir, "Seam carving for content-aware image resizing," *ACM Trans. Graph.*, vol. 26, no. 3, pp.10, 2007

[9] M. Rubinstein, A. Shamir, and S. Avidan, "Improved seam carving for video retargeting," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 16, 2008.

[10] B. Chen, and P. Sen, "Video carving," EUROGRAPHICS, April 2008, Crete, Greece

[11] Z. Li, P. Ishwar, and J. Konrad, "Video Condensation by Ribbon Carving", *IEEE Trans. on Image Processing*, vol.18, no. 11, pp. 2572 – 2583 , 2009

[12] M. Décombas, et al., "Spatio-temporal saliency based on rare model", soumis à *IEEE Proc. Int. Conf. on Image Processing*, 2013

[13] M. Décombas, F. Dufaux, E. Renan, B. Pesquet-Popescu, and F. Capman, "Improved seam carving for semantic video coding",*IEEE Int. Conf. on MultiMedia Signal Processing (MMSP2012)*, 2012, Banff, AB, Canada