Estimation non supervisée de cartes de saillance dynamique dans des vidéos

Léo Maczyta[1] Patrick Bouthemy[1] Olivier Le Meur[2]

[1] Inria, Centre Rennes - Bretagne Atlantique
[2] Univ Rennes, CNRS, IRISA Rennes

leo.maczyta@inria.fr

Résumé

Cet article traite du problème de l'estimation de la saillance du mouvement dans des vidéos. Il consiste à identifier les régions ayant un mouvement se démarquant de son contexte. Nous proposons une nouvelle méthode non supervisée pour calculer des cartes de saillance du mouvement. L'ingrédient central en est l'étape « d'inpainting » du flot optique. Dans un premier temps, des régions candidates sont proposées à partir des frontières du flot optique. Le flot résiduel dans ces régions est obtenu comme la différence entre le flot optique et le flot reconstruit à partir du mouvement environnant. Ce flot résiduel fournit l'information nécessaire pour estimer des cartes de saillance. Notre méthode s'appuie exclusivement sur le mouvement, ce qui la rend flexible et générale. Des résultats expérimentaux sur la base de données DAVIS montrent que la méthode se compare favorablement à des méthodes existantes d'estimation de la saillance dans des vidéos.

Mots Clef

Saillance du mouvement, inpainting du flot optique, analyse de vidéos

Abstract

The paper addresses the problem of motion saliency in videos, that is, identifying regions that undergo motion departing from its context. We propose a new unsupervised paradigm to compute motion saliency maps. The key ingredient is the flow inpainting stage. Candidate regions are determined from the optical flow boundaries. The residual flow in these regions is given by the difference between the optical flow and the flow inpainted from the surrounding areas. It provides the cue for motion saliency. The method is flexible and general by relying on motion information only. Experimental results on the DAVIS 2016 benchmark demonstrate that the method compares favourably with state-of-the-art video saliency methods.

Keywords

Motion saliency, optical flow inpainting, video analysis

1 Introduction

L'estimation de cartes de saillance du mouvement consiste à localiser dans chaque image d'une vidéo la saillance induite par le mouvement. Plus précisément, dans chaque image de la vidéo, les régions dont le mouvement se démarquera suffisamment de leur mouvement environnant seront considérées comme saillantes. L'estimation de la saillance du mouvement peut être utile pour des applications variées, notamment la robotique mobile ou les véhicules autonomes, la génération d'alertes pour la vidéo-surveillance, ou encore l'identification de portions temporelles d'intérêt pour l'interprétation de vidéos.

Les vocables généralement utilisés dans la bibliographie sont plutôt ceux de saillance spatio-temporelle ou de saillance dynamique. Ils englobent de fait une acceptation plus large de la saillance incluant l'apparence. De plus, les principales méthodes d'estimation de saillance dans des vidéos cherchent soit à appréhender l'attention visuelle, soit à mettre en évidence l'objet mobile au premier plan de la scène. Contrairement aux méthodes existantes, nous calculons la saillance du mouvement, et ce, par définition, uniquement à partir de l'information de mouvement. Nous n'exploitons pas d'indices liés à l'apparence, sur laquelle nous ne faisons aucune hypothèse particulière. Notre méthode est d'un côté plus focalisée (sur le mouvement), mais d'un autre plus générale (car dans des situations assez courantes, la saillance peut ne ré-

sulter que du mouvement).

De plus, notre méthode ne requiert aucune étape d'apprentissage, qu'elle soit supervisée ou non. Notre principale contribution consiste à introduire « l'inpainting » du flot optique pour concevoir une solution originale et efficace au problème de l'estimation de la saillance du mouvement. Le recours au flot optique est naturel car il fournit une information dense et la plus complète possible sur le mouvement présent dans la vidéo. En pratique, nous pouvons utiliser n'importe quelle méthode de flot optique, du moment qu'elle soit suffisamment précise, en particulier quant à la préservation des frontières de mouvement, et efficace en termes de temps de calcul. Le flot optique peut indifféremment provenir de méthodes classiques, par exemple de type variationnel, ou de méthodes s'appuyant sur des réseaux de neurones profonds, comme le sont aujourd'hui la très grande majorité des méthodes les plus performantes. Dans les deux cas, nous utiliserons telles quelles ces méthodes (les codes disponibles), et nous récupérerons le flot optique produit comme entrée de notre méthode de saillance du mouvement. Que ces méthodes de flot optique aient pu elles-mêmes impliquer ou non une phase d'apprentissage dans leur construction propre, n'a pas d'incidence sur la nature complètement non supervisée et sans apprentissage de notre méthode de saillance du mouvement.

L'estimation de la saillance dans des vidéos a été d'abord étudiée en tant qu'extension de la saillance d'images, avec pour objectif d'extraire les objets saillants dans la vidéo. Pour le traitement des vidéos, l'information temporelle, et en particulier le mouvement, devient prépondérante pour prédire la saillance. Dans [20], Wang et al. s'appuient sur des informations de contours et de contraste dans les images, ainsi que sur le mouvement pour prédire la saillance dans des vidéos. Dans [10], Le et Sugimoto proposent un modèle « centre-pourtour » qui inclut une segmentation hiérarchique. Dans [7], Karimi et al. exploitent des indices spatio-temporels et représentent les vidéos comme des graphes spatio-temporels, avec pour objectif de minimiser une fonction globale.

Le mouvement apparent dans chaque image est fortement influencé par le mouvement de la caméra. Un premier groupe de méthodes combine directement les informations spatiales et temporelles sans chercher à compenser le mouvement de la caméra, dont par exemple [3, 8, 12]. Un second groupe de méthodes compense explicitement le mouvement de la caméra, comme [11] ou encore [5].

Récemment, des méthodes s'appuyant sur l'apprentissage profond ont été proposées pour estimer la saillance dans des vidéos. Dans [21], Wang et al. retiennent un réseau convolutionnel (CNN) exploitant explicitement les dimensions spatiales et temporelle, sans toutefois calculer le flot optique. Dans [9], Le et Sugimoto utilisent des caractéristiques spatio-temporelles profondes pour prédire la saillance dynamique dans des vidéos. Ils étendent les champs aléatoires conditionnels au domaine temporel, et utilisent une stratégie de segmentation multi-échelle. Dans [19], Wang et al. considèrent la saillance du mouvement comme un *a priori* pour la tâche de segmentation d'objets dans des vidéos, en utilisant les contours spatiaux et les frontières de mouvement comme caractéristiques.

Les méthodes évoquées ci-dessus s'intéressent principalement au problème de la saillance vidéo définie à partir d'une notion *d'objet*. Ces méthodes ont pour but l'extraction d'objets situés au premier plan et se démarquant de leur contexte par leur apparence et leur mouvement.

Nous nous intéressons plus particulièrement au problème de l'estimation de la saillance de mouvement, qui se focalise sur l'identification de mouvements remarquables. Les éléments repérables peuvent en effet n'être dus qu'au mouvement, comme dans la détection d'anomalies dans des foules [15] (une personne se déplaçant différemment de la foule environnante, ou de façon similaire un animal dans un groupe, une voiture dans la circulation, ou encore une cellule dans un tissu). De plus, l'apparence peut n'apporter que peu d'informations dans certains types d'imageries, comme par exemple pour des vidéos prises par des caméras infrarouges ou pour de la microscopie par fluorescence.

Le reste de l'article est organisé comme suit. Dans la section 2, nous présentons notre méthode d'estimation de la saillance du mouvement. La section 3 montre des résultats comparatifs avec des méthodes de l'état de l'art pour la saillance vidéo. Enfin, la section 4 conclut l'article.

2 Estimation de la saillance de mouvement

Comme précisé dans l'introduction, nous estimons des cartes de saillance du mouvement uniquement à partir d'informations extraites du flot optique. Le flot

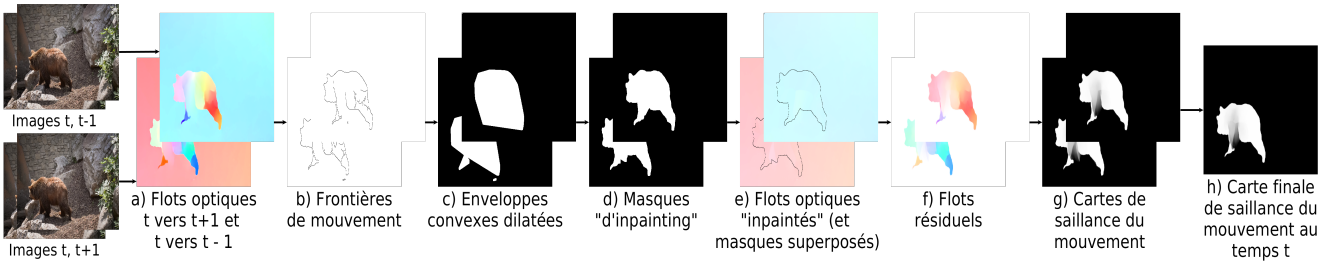

FIGURE 1 – Schéma général de notre méthode pour l'estimation de cartes de saillance du mouvement, avec les deux traitements dans le sens du temps et dans le sens rétrograde

optique est supposé être suffisamment distinct dans des régions saillantes. Nous comparons le flot optique dans une région donnée, susceptible de contenir un élément saillant, au flot qui aurait été induit dans cette zone par le mouvement environnant. Toute méthode de calcul du flot optique peut être utilisée pour estimer le flot sur l'ensemble de l'image. Le flot induit par le mouvement environnant n'est pas directement disponible, puisqu'il n'est pas observé. Il peut cependant être prédit par une méthode « d'inpainting » du flot optique. C'est ce point qui fait l'originalité de notre méthode d'estimation de la saillance du mouvement. Notre méthode se divise en deux étapes. Nous commençons par extraire des régions saillantes candidates dans lesquelles nous comparons le flot reconstruit par « inpainting » avec le flot original. Une différence élevée entre les deux flots est ensuite interprétée comme un indicateur de la présence de saillance du mouvement. De plus, nous combinons un traitement temporel dans le sens du temps et dans le sens réciproque. L'architecture globale de la méthode est fournie à la figure 1.

2.1 Extraction des régions candidates

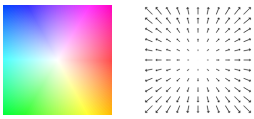

FIGURE 2 – Code couleur (à gauche) pour le flot optique correspondant (à droite).

Tout d'abord, nous extrayons les masques des régions à « inpainter ». Nous exploitons le flot optique calculé sur toute l'image, ainsi que ses discontinuités. En effet, le contour des éléments dynamiquement saillants est censé correspondre aux frontières de mouvement, comme leur mouvement doit se démarquer du mouvement environnant. Le mouvement environnant correspondra généralement au mouvement apparent de l'arrière-plan dû au mouvement de la caméra. Il sera aussi désigné mouvement global dans la suite.

Pour l'extraction des frontières de mouvement, un seuillage sur la norme du gradient des vecteurs de vitesse pourrait être directement appliqué. Cependant, cela risquerait de produire des contours trop bruités. À la place, nous choisissons de nous appuyer sur la méthode classique d'extraction de contour proposée par Canny [2]. Pour cela, nous convertissons le flot optique en sa représentation HSV, qui est prise en entrée de l'algorithme de Canny. La représentation HSV est communément utilisée pour la visualisation du flot optique, avec la teinte qui représente la direction du mouvement et la saturation son amplitude (voir figure 2).

Nous construisons ensuite des masques à partir de ces contours potentiellement fragmentés, comme illustré à la figure 1b)-d). Les contours sont tout d'abord regroupés en parties connexes. L'enveloppe convexe de chaque partie connexe est calculée et dilatée avec un noyau 5x5. Les masques sont obtenus en prenant l'union des enveloppes convexes dilatées se chevauchant. Par construction, les masques tendent à être plus larges que les régions saillantes qu'ils sont censés recouvrir. Cette propriété est souhaitable pour « l'inpainting », puisque celui-ci ne devrait utiliser que le mouvement global pour la reconstruction. Néanmoins, un masque trop grossier est susceptible de diminuer la précision de la reconstruction du flot, en particulier pour les zones saillantes qui sont non convexes (voir figure 1c)). Les masques sont donc affinés en appliquant l'algorithme du GrabCut [16] sur la représentation HSV du flot optique. Enfin, pour éviter d'avoir de petites erreurs de localisation qui conduiraient à inclure des pixels saillants en bordure du masque « d'inpainting », une dilatation avec un noyau 5x5 est de nouveau appliquée à chaque masque résultant.

2.2 Inpainting du flot optique

Nous disposons d'un ensemble E de masques pour « l'inpainting » dans le domaine de l'image Ω. L'objectif est maintenant de reconstruire le flot dans ces masques à partir du mouvement environnant. Nous avons pour ce faire testé trois méthodes « d'inpainting » : deux méthodes utilisant des EDP [18, 1] et une méthode paramétrique. Les méthodes s'appuyant sur des EDP semblent particulièrement bien adaptées au problème étant donné la régularité du champ des vitesses à reconstruire.

Nous appliquons au flot optique la méthode « d'inpainting » exploitant le *fast marching* de [18], de la même façon que ce qui a été fait par [17] pour l'objectif de reconstruction de vidéos. Nous étendons, de façon similaire, au flot optique la méthode « d'inpainting » de [1] qui s'inspire des équations de Navier-Stokes. Nous utilisons la représentation réelle des deux composantes des vecteurs de vitesse $\{\boldsymbol{\omega}(p), p \in \Omega\}$ avec $\boldsymbol{\omega}(p) \in \mathbb{R}^2$. Les deux composantes du vecteur $\boldsymbol{\omega}(p)$ sont reconstruites séparément. Finalement, nous avons développé une alternative paramétrique. Nous supposons que le mouvement global peut être approché par un modèle paramétrique affine. Celui-ci est estimé par la méthode multi-échelle robuste Motion2D [13]. Le flot reconstruit est ensuite simplement obtenu comme le flot fourni par le modèle affine sur les masques. Ces trois variantes sont respectivement notées MSI-fm, MSI-ns and MSI-pm (MSI signifiant *Motion Saliency Inpainting*).

2.3 Calcul des cartes de saillance de mouvement

Le flot résiduel, qui correspond à la différence entre le flot optique et le flot reconstruit, est calculé sur les masques $r, r \in E$. À partir du flot résiduel, l'objectif est d'obtenir une carte de saillance du mouvement g, à valeurs scalaires dans $[0, 1]$. g est défini de la façon suivante :

$$g(p) = 1 - \exp(-\lambda ||\boldsymbol{\omega}_{inp}(p) - \boldsymbol{\omega}(p)||_2), \qquad (1)$$

où p est un pixel d'un masque $r \in E$, $\boldsymbol{\omega}$ est le flot optique, $\boldsymbol{\omega}_{inp}$ est le flot reconstruit, et λ module le score de saillance . Pour les pixels p hors des masques d'inpainting, $g(p) = 0$. La fonction g traduit l'idée que les mouvements résiduels non nuls dénotent la présence d'éléments ayant un mouvement potentiellement saillant. Le paramètre λ nous permet d'établir un compromis entre la robustesse au bruit et la capacité à souligner des mouvements même faiblement saillants.

Notons que si nous étions intéressés par une segmentation explicite par le mouvement, c'est-à-dire produire une carte binaire, il suffirait de fixer λ à une valeur élevée. En effet, en appliquant un seuil τ à $g(p)$, nous pouvons déduire de (1) que p appartiendra aux zones de mouvement segmentées si :

$$||\boldsymbol{\omega}_{inp}(p) - \boldsymbol{\omega}(p)||_2 \geqslant -\frac{ln(1-\tau)}{\lambda}. \qquad (2)$$

En fixant τ arbitrairement à $\frac{1}{2}$ (le milieu de $[0, 1]$), la décision ne dépend plus que de λ. Les pixels pour lesquels le flot résiduel a une amplitude supérieure à $\frac{ln(2)}{\lambda}$ seront extraits. Ainsi, notre méthode est flexible, comme le montre la possibilité de passer du problème de l'estimation de la saillance du mouvement à un problème de segmentation, en jouant simplement sur le paramètre λ.

Pour finir, nous proposons d'aller plus loin dans la prise en compte de la dimension temporelle pour réduire le nombre de faux positifs, en particulier proche des frontières de mouvement. Pour ce faire, nous nous appuyons sur un traitement bidirectionnel (voir Fig. 1a)-g)). Le traitement décrit ci-dessus est appliqué deux fois en parallèle, sur les paires d'images $I(t), I(t-1)$ et $I(t), I(t+1)$. On obtient ainsi deux cartes de saillance, qui sont ensuite combinées en prenant la saillance minimale prédite pour chaque pixel. Les résultats expérimentaux présentés pour MSI-fm, MSI-ns, MSI-pm et la méthode NM qui sera détaillée en Section 3.3 incluront ce traitement bidirectionnel.

3 Résultats expérimentaux

3.1 Protocole expérimental

Le flot optique est calculé avec l'algorithme FlowNet 2.0 [6]. Cet algorithme, rapide et performant, produit des frontières de mouvement bien marquées. Ce point est important pour l'étape d'extraction des masques « d'inpainting ».

Pour l'ensemble des expériences, les paramètres sont fixés de la façon suivante. Le détecteur de contour de Canny est appliqué à l'image lissée avec un filtre gaussien de déviation standard $\sigma = 5$. Les deux seuils pour le détecteur de Canny sont fixés à 20 et 60. Le ratio de 3 entre les deux valeurs est choisi suivant les recommandations de [2]. Pour l'algorithme « d'inpainting », un rayon de 5 pixels autour de la région à « inpainter » est utilisé. Notons que ce rayon de 5

pixels ainsi que le choix d'un noyau 5x5 pour les opérations de dilatation ont été utilisés pour des dimensions d'images variables. Les images de DAVIS ont une dimension de 854x480, et les exemples additionnels qui seront présentés figure 3 ont des dimensions de 720x720 pour l'exemple synthétique, et de 352x288 pour l'exemple de vidéo infrarouge. Enfin, le paramètre λ pour le calcul des cartes de saillance a été fixé à $\frac{3}{2}$.

Il n'existe pas de base de données véritablement dédiée à l'estimation de la saillance du mouvement. Nous avons utilisé DAVIS 2016 pour l'évaluation de notre méthode. Cette base de vidéos a été proposée dans [14] pour la tâche de segmentation d'objets dans des vidéos (VOS). Elle a également été récemment utilisée par des méthodes d'estimation de cartes de saillance dans des vidéos, comme par exemple [9, 21]. Pour la tâche VOS, l'objet à segmenter est un objet au premier plan de la vidéo dont le mouvement est fortement distinct comparé au reste de la scène. Cette caractéristique rend cette base de données utilisable pour l'estimation de la saillance du mouvement, bien que l'apparence intervienne également dans la définition de la vérité-terrain. Plus précisément, cette dernière contient tout l'objet, même si certaines de ses parties ne bougent pas.

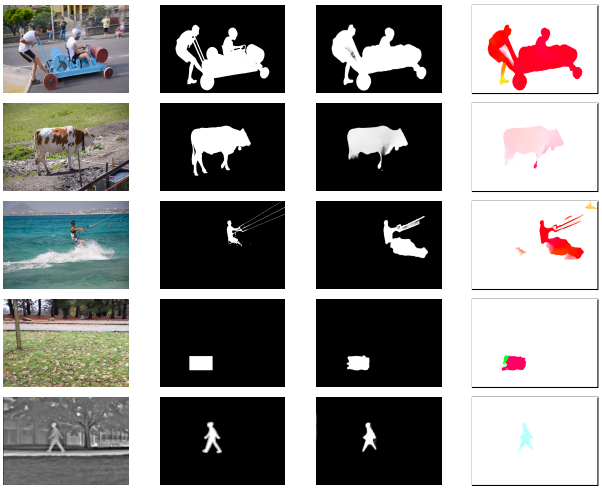

FIGURE 3 – De gauche à droite : a) une image d'une vidéo, b) la vérité-terrain binaire, c) la carte de saillance du mouvement prédite par notre méthode MSI-ns, et d) le flot résiduel entre les images $I(t)$ et $I(t+1)$. Le flot résiduel est représenté avec le code couleur fourni en figure 2.

3.2 Évaluation qualitative

Le tableau 1 montre que la méthode MSI-ns est la meilleure des trois variantes proposées, c'est donc celle-ci qui sera prise dans cette évaluation qualitative. La figure 3 montre les résultats de notre méthode pour de haut en bas des images des vidéos *soapbox, cows* et *kite-surf* de DAVIS 2016, ainsi que pour deux autres types de vidéos. Dans le quatrième exemple (vidéo de la pelouse), une région rectangulaire de l'herbe a été artificiellement déplacée dans l'image comme indiqué par la vérité-terrain. Cet exemple illustre le comportement de la méthode lorsque la seule information discriminative est liée au mouvement. La cinquième image vient de la vidéo *park* de la base de données changedetection.net [4]. Cette vidéo a été acquise avec une caméra infrarouge, et constitue ainsi un exemple où l'apparence n'apporte qu'une aide limitée.

La carte de saillance du mouvement et le flot résiduel sont présentés respectivement en figure 3c) et 3d). En effet, bien que le flot résiduel soit obtenu lors d'une étape intermédiaire de notre méthode, il n'en est pas moins exploitable en tant que tel. Le flot résiduel fournit des informations additionnelles utiles sur la direction et l'amplitude des mouvements saillants dans la scène. Il pourrait en ce sens être vu comme une carte de saillance augmentée.

Pour l'exemple de la vidéo *soapbox*, l'élément saillant avec un mouvement clairement distinctif a été extrait de façon satisfaisante. Le second exemple *cows* illustre quant à lui un comportement intéressant. La vache est globalement en mouvement, à l'exception de ses pattes qui sont statiques par intermittence. Ceci illustre la différence entre d'une part la tâche de segmentation d'objet dans les vidéos, pour laquelle l'objectif serait de segmenter la vache en entier, et d'autre part la tâche d'estimation de la saillance du mouvement, pour laquelle les éléments d'intérêt sont les éléments avec un mouvement saillant. De façon cohérente avec ce dernier objectif, notre méthode n'inclut pas les pattes statiques dans la carte de saillance.

Pour l'exemple *kite-surf*, l'écume a un fort mouvement, ce qui la fait naturellement apparaître comme saillante au sens du mouvement, tandis que pour la segmentation d'objet vidéo, le surfeur est le seul objet au premier plan à segmenter.

Pour l'exemple de la pelouse, la région carrée est facilement identifiable lors de la visualisation de la vidéo, mais elle est bien plus difficile à localiser sur une

Méthode	STCRF [9]	MSI-ns	MSI-pm	MSI-fm	VSFCN [21]	RST [10]	LGFOGR [20]	SAG [19]	NM
MAE ↓	**0,033**	0,043	0,044	0,045	0,055	0,077	0,102	0,103	0,453
F-Adap ↑	**0,803**	0,735	0,724	0,716	0,698	0,627	0,537	0,494	0,367
F-Max ↑	**0,816**	0,751	0,750	0,747	0,745	0,645	0,601	0,548	0,612
Apparence	*Oui*	*Non*	*Non*	*Non*	*Oui*	*Oui*	*Oui*	*Oui*	*Non*
Mouvement	*Oui*	*Oui*	*Oui*	*Oui*	*Oui*	*Oui*	*Oui*	*Oui*	*Oui*
Supervisé	*Oui*	*Non*	*Non*	*Non*	*Oui*	*Non*	*Non*	*Non*	*Non*

TABLE 1 – Comparaison avec des méthodes de l'état de l'art de l'estimation de cartes de saillance sur l'ensemble de test de DAVIS 2016. Les meilleures performances sont en gras, les secondes meilleures sont soulignées. Le sens de la flèche, à la droite du critère, indique si les meilleures performances sont obtenues pour des valeurs faibles ou élevées. Il est aussi indiqué si la méthode s'appuie sur des indices liés à l'apparence, des indices liés au mouvement, et si la méthode est supervisée.

seule image statique. Notre méthode, qui exploite le flot optique, parvient à identifier la région saillante en mouvement. Finalement, pour l'exemple de *park* filmé avec une caméra infrarouge où l'apparence est moins discriminative, notre méthode produit une carte de saillance du mouvement correcte.

3.3 Comparaison quantitative

Pour l'évaluation quantitative, nous avons recours à l'erreur moyenne absolue (MAE), ainsi qu'aux métriques F-Adap et F-Max, que nous calculons de la même façon que dans [9]. La MAE est une évaluation au niveau des pixels de la carte de saillance g, qui est comparée à une vérité-terrain binaire. F-Adap et F-Max sont dérivées d'une F-Mesure pondérée, pour laquelle le poids β^2 est fixé à 0,3, comme indiqué dans [9] :

$$F_\beta = \frac{(1+\beta^2) Precision \times Recall}{\beta^2 \times Precision + Recall} \qquad (3)$$

F-Adap et F-Max nécessitent de seuiller la carte de saillance. F-Adap implique un seuillage adaptatif sur chaque carte de saillance, ce seuillage étant construit sur la moyenne et l'écart type des valeurs de saillance de chaque carte. F-Max correspond simplement au maximum des F-Mesures obtenues pour des seuils variant sur [0, 255].

Nous présentons ici une méthode naïve de saillance de mouvement (notée NM) pour mieux comprendre l'intérêt des composantes principales de notre méthode. Dans un premier temps, le mouvement dominant (ou mouvement global) est calculé sur l'ensemble de l'image. Pour cela, nous estimons un modèle affine de mouvement avec l'algorithme robuste multi-échelle Motion2D [13]. Aucun masque « d'inpainting » n'est extrait, ce qui représente la différence principale avec nos autres méthodes. Le flot résiduel utilisé pour le calcul de la carte de saillance est directement obtenu comme la différence, sur l'image entière, entre le flot optique et le flot dominant correspondant au modèle paramétrique estimé. Comme le montre le tableau 1, la méthode NM présente des performances très inférieures aux autres méthodes. Ceci démontre l'importance de l'approche par « inpainting » du flot pour l'estimation de la saillance de mouvement.

Le tableau 1 contient par ailleurs des résultats comparatifs des trois variantes de notre méthode, MSI-ns, MSI-pm et MSI-fm, avec des méthodes de l'état de l'art d'estimation de cartes de saillance dans des vidéos : LGFOGR [20], SAG [19], RST [10], STCRF [9] et VSFCN [21]. Les résultats pour ces méthodes sont rapportés par [9], à l'exception de [21] pour laquelle nous avons repris les cartes de saillance fournies par les auteurs pour le calcul des métriques. Nous avons mené cette expérience sur l'ensemble de test de DAVIS 2016, qui contient vingt vidéos. L'évaluation quantitative sur DAVIS 2016 est informative, mais est susceptible de présenter un (léger) biais. La vérité-terrain disponible sur DAVIS 2016 peut ne pas être exactement appropriée pour l'estimation de la saillance du mouvement comme illustré en figure 3 et commenté en section 3.2, étant donné qu'elle est centrée sur la segmentation d'objets et est purement binaire.

Notre méthode MSI-ns obtient des résultats satisfaisants et cohérents, étant placée seconde pour les trois métriques. Les deux autres variantes, MSI-pm et MSI-fm, sont classées respectivement troisième et quatrième, tout en restant très proches des performances de MSI-ns. Rappelons que nos résultats sont obte-

nus sans apprentissage sur la saillance et sans utiliser d'indications de saillance liées à l'apparence, contrairement à [9] qui obtient les meilleurs résultats. Les variantes de notre méthode qui s'appuient respectivement sur le modèle paramétrique (MSI-pm) et sur la diffusion (MSI-ns et MSI-fm) ont des performances similaires sur DAVIS 2016. Cependant, les secondes devraient être plus facilement généralisables, étant donné que le mouvement de l'arrière-plan ne peut pas toujours être approché par un unique modèle paramétrique. Si le modèle paramétrique est restreint à la zone environnante, la question de la spécification de cette zone va alors se poser.
En ce qui concerne le temps de calcul, la méthode MSI-ns calcule la carte de saillance sur une image 854x480 en 1.2 secondes sur un processeur à 2,6 GHz. Notre code est écrit en Python et peut être encore optimisé. En particulier, les deux traitements dans les deux sens temporels pourraient être parallélisés.

4 Conclusion

Nous avons défini une nouvelle méthode pour estimer des cartes de saillance du mouvement dans des vidéos, qui repose essentiellement sur la reconstruction du flot optique. Cette méthode produit des cartes de saillance valuées qui indiquent la présence de saillance du mouvement dans les vidéos. Nous avons testé notre méthode sur la base de vidéos DAVIS 2016 et avons obtenu des résultats satisfaisants, tout en utilisant uniquement l'information du mouvement et sans introduire d'étape d'apprentissage. Ces caractéristiques rendent notre méthode générale. De plus, le flot résiduel estimé fournit par lui-même une information augmentée sur la saillance du mouvement, qui pourrait être exploitée directement. Notre méthode s'appuie sur trois images successives. Des travaux futurs pourront étudier une meilleure prise en compte de la dimension temporelle.

Remerciements

Ces travaux ont été partiellement financés par la DGA et la Région Bretagne par des co-financements de la thèse de Léo Maczyta.

Références

[1] M. Bertalmio, A. L. Bertozzi, and G. Sapiro. Navier-stokes, fluid dynamics, and image and video inpainting. In *CVPR*, 2001.

[2] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6) :679–698, Nov 1986.

[3] Y. Fang, Z. Wang, W. Lin, and Z. Fang. Video saliency incorporating spatiotemporal cues and uncertainty weighting. *IEEE Transactions on Image Processing*, 23(9) :3910–3921, Sept 2014.

[4] N. Goyette, P. Jodoin, F. Porikli, J. Konrad, and P. Ishwar. Changedetection.net : A new change detection benchmark dataset. In *CVPRW*, 2012.

[5] C. R. Huang, Y. J. Chang, Z. X. Yang, and Y. Y. Lin. Video saliency map detection by dominant camera motion removal. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(8) :1336–1349, Aug 2014.

[6] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox. Flownet 2.0 : Evolution of optical flow estimation with deep networks. In *CVPR*, 2017.

[7] A. H. Karimi, M. J. Shafiee, C. Scharfenberger, I. BenDaya, S. Haider, N. Talukdar, D. A. Clausi, and A. Wong. Spatio-temporal saliency detection using abstracted fully-connected graphical models. In *ICIP*, 2016.

[8] W. Kim and C. Kim. Spatiotemporal saliency detection using textural contrast and its applications. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(4) :646–659, April 2014.

[9] T. Le and A. Sugimoto. Video salient object detection using spatiotemporal deep features. *IEEE Transactions on Image Processing*, 27(10) :5002–5015, Oct 2018.

[10] T.-N. Le and A. Sugimoto. Contrast based hierarchical spatial-temporal saliency for video. In *Image and Video Technology*, pages 734–748, 2016.

[11] O. Le Meur, P. Le Callet, and D. Barba. Predicting visual fixations on video based on low-level visual features. *Vision Research*, 47(19) :2483–2498, 2007.

[12] D. Mahapatra, S. O. Gilani, and M. K. Saini. Coherency based spatio-temporal saliency detection for video object segmentation. *IEEE Journal of Selected Topics in Signal Processing*, 8(3) :454–462, June 2014.

[13] J.-M. Odobez and P. Bouthemy. Robust multiresolution estimation of parametric motion models. *Journal of Visual Communication and Image Representation*, 6(4) :348 – 365, 1995.

[14] F. Perazzi, J. Pont-Tuset, B. McWilliams, L. V. Gool, M. Gross, and A. Sorkine-Hornung. A benchmark dataset and evaluation methodology for video object segmentation. In *CVPR*, 2016.

[15] J.-M. Pérez-Rúa, A. Basset, and P. Bouthemy. Detection and localization of anomalous motion in video sequences from local histograms of labeled affine flows. *Frontiers in ICT, Computer Image Analysis*, 2017.

[16] C. Rother, V. Kolmogorov, and A. Blake. Grabcut -interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics (SIGGRAPH)*, pages 309–314, August 2004.

[17] M. Strobel, J. Diebold, and D. Cremers. Flow and color inpainting for video completion. In *CVPR*, 2014.

[18] A. Telea. An image inpainting technique based on the fast marching method. *Journal of Graphics Tools*, 9 :23–34, Jan 2004.

[19] W. Wang, J. Shen, and F. Porikli. Saliency-aware geodesic video object segmentation. In *CVPR*, 2015.

[20] W. Wang, J. Shen, and L. Shao. Consistent video saliency using local gradient flow optimization and global refinement. *IEEE Transactions on Image Processing*, 24(11) :4185–4196, Nov 2015.

[21] W. Wang, J. Shen, and L. Shao. Video salient object detection via fully convolutional networks. *IEEE Transactions on Image Processing*, 27(1) :38–49, Jan 2018.