HAL
open science

Détection d'événements rares par modélisation de la signature du flot optique

Dieudonné Fabrice Atrevi, Damien Vivet, Bruno Emile

To cite this version:

Dieudonné Fabrice Atrevi, Damien Vivet, Bruno Emile. Détection d'événements rares par modélisation de la signature du flot optique. GRETSI, Aug 2019, Lille, France. hal-02277109

HAL Id: hal-02277109

https://hal.science/hal-02277109

Submitted on 3 Sep 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Détection d'événements rares par modélisation de la signature du flot optique

Dieudonné Fabrice ATREVI[1], Damien VIVET[2], Bruno EMILE[1]

[1]Université d'Orléans, INSA Centre Val de Loire, PRISME EA 4229, F45072, Orléans, France

[2]ISAE-Supaero, Université de Toulouse, Toulouse, France

fabrice.atrevi@univ-orleans.fr, damien.vivet@isae.fr
bruno.emile@univ-orleans.fr

Résumé – Nous présentons dans cet article une nouvelle méthode de détection d'événements rares basée sur l'analyse des mouvements saillants dans une scène. La méthode vise à localiser automatiquement toutes les régions d'une scène où se déroulent des événements rares. Un événement rare s'oppose aux événements fréquents en ce sens qu'on y intégre tout événement nouveau dans la scène. Elle exploite les propriétés de la transformée en cosinus discrète (TCD) qui, appliquée à une image, permet d'y détecter des irrégularités spatiales. En appliquant successivement la transformée et son inverse aux données de magnitude et d'orientation du flot optique, notre méthode permet de localiser tous les mouvements irréguliers. Une modélisation de ces mouvements, grâce à la version "one class" de l'algorithme SVM, permet d'identifier sur les événements rares parmi ceux fréquents. La méthode a été testée sur la base publique UCSD [1] et présente des résultats satisfaisants.

Abstract – We present, in this paper, a new method for rare events détection based on salient motion analysis. The method aims to automatically locate all regions of a scene where rare events occur. A rare event is an event that not occurs frequently in the scene. It exploits properties of the discrete cosine transform (DCT) that can detect spatial irregularities in images. By performing the DCT and its inverse on the magnitude and orientation of the optical flow, our method makes it possible to locate any salient motion. Modeling of the salient motions thanks to the "one class" version of the support vector machine (SVM) allow the identification of rare events between salient motion. The method has been tested on the UCSD dataset and shows that it's promising.

1 Introduction

La détection automatique de mouvements saillants dans une scène est une tâche importante pour de nombreuses applications en vision par ordinateur, tels les systèmes de vidéo surveillance. Pour de tel système, ces méthodes visent à réduire l'effort cognitif des surveillants en proposant des zones d'activités irrégulières de la scène car comme le soulignent Green et al. [2], l'attention des opérateurs de vidéosurveillance tend à se dégrader après une longue période de surveillance. Les zones mises en évidence peuvent être le siège du déroulement d'événements rares. La localisation des événements rares est donc une couche supérieure dans le sytème. Elle implique une classification des mouvements saillants en identifant clairement parmi les mouvements saillants à un instant t, ceux qui n'ont jamais été observé auparavant. Dans ce papier, nous proposons une méthode qui part de la détection des mouvements saillants dans une scène à un instant précis pour déboucher sur la détection des événements rares. La mise en place d'un tel système nécessite de relever de nombreux défis, en particulier dans le domaine de la vision par ordinateur.

De nombreuses approches ont été proposées dans la littérature pour efficacement détecter et localiser les événements rares dans des vidéos de scène naturelle. Dans leur revue, Thida et al. [3] ont identifiés deux catégories d'approches : les approches de modélisation macroscopiques et celles microscopiques. Notre approche appartient à la deuxième catégorie. Les récentes approches de la littérature essaient d'améliorer les performances de la localisation des événements plutôt que l'identification des séquences d'images contenant d'éventuels événements rares. Shi et al. dans [4] ont proposé de modéliser les événements normaux en utilisant comme caractéristique l'histogramme multi-échelle du flot optique. À des fins de localisation, ils ont utilisé un algorithme de saillance visuelle spatio-temporel pour sélectionner les régions d'extraction de ces caractéristiques. Lei et al. [5] ont proposé un nouveau descripteur dénommé "Spatial Associated Multi-scale Histogram of Optical Flow" (SA-MHOF). Le descripteur a pour objectif de résoudre le problème de déformation perspective lié aux objets éloignés de la caméra en intégrant la position de l'extraction du descripteur dans la scène lors de l'étape de calcul du flot optique. De nombreux autres articles [6, 7, 8] ont proposé différents descripteurs.

Notre méthode s'inspire de la méthode de détection d'objets saillants de Hou et al. [9] basée sur l'analyse d'image avec la transformée en cosinus discrète. Elle utilise les informations d'orientation et de vitesse du flot optique pour d'une part détecter les mouvements saillants puis identifie ceux correspondant à un événement rare.

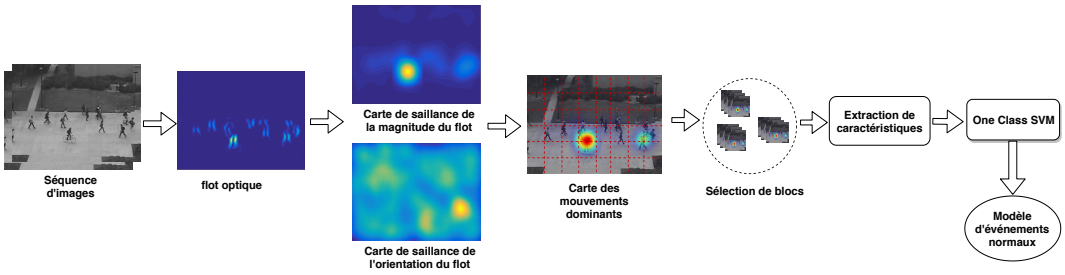

FIGURE 1 – Méthodologie de l'approche proposée

2 Approche proposée

L'approche peut être scindée en quatre étapes complémentaires, comme indiqué sur la Fig 1. Elle prend en entrée deux images consécutives et donne en sortie la carte des événements rares qui met en surbrillance les différentes régions d'intérêt de la scène. Les différentes étapes sont décrites ci-dessous :

— On commence par calculer le flot optique de deux images consécutives afin d'estimer le champ de mouvement dans la scène.

— La deuxième étape consiste à calculer la signature de l'orientation et de la magnitude du champ de mouvement.

— A partir des signatures de l'orientation et de la magnitude, nous estimons la distribution des mouvements en générant une carte de saillance. Cette carte est par la suite divisée en des blocs parmi lesquels seuls ceux suffisamment saillants sont retenus pour l'étape suivante.

— La dernière étape consiste à construire un vecteur de caractéristiques pour chaque bloc retenu. Ce vecteur est obtenu par vectorisation de la matrice de score de saillance de tous les pixels du bloc. Pour la phase d'apprentissage, l'algorithme SVM one-class est utilisé pour construire un modèle d'événements à partir uniquement d'événements fréquents.

2.1 Estimation du champ de mouvement

On désigne par champ de mouvement l'ensemble des déplacements dans la scène. Il fournit les informations telles que la direction et la vitesse de chaque pixel dans l'image et peut être estimé avec de nombreuses approches. Dans cet article, nous avons utilisé l'algorithme de Horn et al. [10]. Soit $I_{x,y,t}$ la valeur d'intensité du pixel (x, y) à l'instant t. Les données d'orientation $\varphi_{x,y,t}$ et de magnitude $r_{x,y,t}$ du vecteur de flot associé à chaque pixel sont exprimées suivant la formulation suivante :

$$\varphi_{x,y,t} = arctan\left(\frac{V_x}{V_y}\right) \tag{1}$$

et

$$r_{x,y,t} = \sqrt{V_x^2 + V_y^2} \tag{2}$$

où V_x et V_y sont les composantes horizontale et verticale du vecteur de flot.

Le but de notre méthode étant de détecter les événements rares en fonction des mouvements dans le scène, nous émettons l'hypothèse qu'une fusion des données d'orientation et de magnitude apporterait plus d'information. En effet, les événements rares peuvent être causés soit par l'orientation du mouvement ou soit par sa vitesse ou par les deux à la fois. Prenons l'exemple du mouvement saillant de la figure 1. Dans ce cas, l'événement est généré par le mouvement saillant du cycliste dont la vitesse est plus importante que celle des piétons sur la scène.

2.2 La signature du flot et la génération de la carte de saillance

Dans notre approche, nous proposons d'utiliser la transformée en cosinus discrète (TCD) pour détecter les irrégularités dans les informations de flot optique. Considérons chaque composante du flot optique qui est un mélange entre les mouvements réguliers et irréguliers selon la structure suivante :

$$y = r + i \quad y, r, i \in \mathbb{R}^N \tag{3}$$

r représente les mouvements réguliers et est supposé être représenté dans la base spatiale standard. i représente les mouvements irréguliers et est supposé être représenté dans la base de la TCD. Hou et al. ont démontré dans [11] qu'il est possible d'isoler approximativement l'arrière-plan dans une image en prenant la transformée inverse du signe de la TCD de l'image. A partir de cette hypothèse, nous définissons la signature de flot optique par 4 qui est le sign de la TCD. La signature de flot rejette les informations d'amplitude sur toute la fréquence et est très compacte car est représenté avec un seul bit par composant.

$$FlotSignature(y) = sign(DCT(y)) \tag{4}$$

où y est soit l'information d'orientation ou soit l'information de magnitude du flot optique.

A la suite du calcul de la signature de l'orientation et de la magnitude, les données du flot sont reconstruites à partir de ces signatures via l'intermédiaire de la Transformée Inverse en Cosinus Discrète.

$$\bar{y} = IDCT(FlotSignature(y)) \tag{5}$$

La carte de saillance est ensuite générée en lissant le carré de la carte reconstruite ci-dessus avec la formulation de 6 où g est

un noyau gaussien. Le filtrage est une étape importante car il élimine le bruit introduit par la quantification du signe.

$$S = g * (\bar{y} \circ \bar{y}) \tag{6}$$

Les cartes d'orientation S_φ et de magnitude S_r sont générées en suivant le même processus. La combinaison des deux types d'informations se fait en multipliant S_φ par S_r :

$$S_f = S_\varphi \circ S_r \tag{7}$$

2.3 Extraction de caractéristiques et construction du modèle d'événements

La carte des mouvements saillants S_f est une matrice de score de même dimension que l'image d'entrée. Les pixels ayant comme score 0 appartiennent à des régions de la scène où aucun mouvement saillant ne se produit tandis que les pixels de mouvement saillant ont des valeurs élevées. Il est important de souligner que les mouvements sont considérés saillants relativement aux mouvements dans la scène à l'instant t. Ils ne peuvent donc pas automatiquement être considérés comme générant les événements rares. Le but de cette étape est de construire un modèle d'événements fréquents à partir d'un jeu d'apprentissage qui ne contient que ces événements. Pour ce faire, chaque carte de saillance est divisée en blocs de taille fixe parmi lesquels seuls les blocs dont la moyenne des scores des pixels est supérieure à un seuil seront pris en compte pour la construction du modèle. La valeur de ce seuil est empiriquement fixée à deux fois la moyenne des scores de l'image courante. Cette étape permet de réduire le nombre de blocs nécessaires à la modélisation. Pour chaque bloc, le vecteur caractéristique est obtenu en vectorisant la matrice des scores du bloc :

$$FeatureVect_i = vect(score(block[i])) \tag{8}$$

où $vect$ est la fonction de vectorisation.

Le modèle d'événements est construit avec l'algorithme SVM one-class. Le but de cet algorithme est de trouver une frontière autour de l'ensemble d'apprentissage. Différents types de frontières peuvent être utilisés : linéaire, circulaire, etc. Dans cet article, nous utilisons un noyau linéaire. Pour identifier le bloc d'événements rares, nous utilisons comme score la distance entre le vecteur de caractéristique et l'origine dans l'espace du SVM. Plus la distance est grande, plus la probabilité que le bloc contienne un événement rare est élevée.

3 Résultats Expérimentaux

L'évaluation de notre approche a été faite sur la base UCSD [1]. La base UCSD est composée de deux ensembles de données distincts : Ped 1 qui contient 34 vidéos d'entraînement et 36 vidéos de test et Ped 2 qui contient 16 vidéos d'entraînement et 13 vidéos de test. Pour l'évaluation quantitative, nous avons utilisé l'aire sous la courbe ROC (AUC) associée à l'EER (Equal Error Rate) et au taux de détection RD (Rate Detection) comme mesures de performance. Plus l'AUC et le RD sont élevés et l'EER est faible, meilleure est la méthode. Dans cet article, nous avons effectué des évaluation de deux niveaux : niveau image et niveau pixel. Pour l'évaluation niveau image, c'est l'image dans sa totalité qui est classée soit contenant d'événements fréquents ou soit contenant des événements rares, en prenant le score le plus élevé de ses blocs comme score de classification. Cette évaluation ne permet pas d'apprécier les performances de localisation de la méthode. Ainsi, pour y rémédier, l'évaluation niveau pixel est approprié. Pour cette évaluation, une image est classée comme contenant des événements rares si les événements rares détectés couvrent au moins 40% de la vérité terrain.

TABLE 1 – Résultats d'évaluation sur la base Ped 1

	Niveau d'évaluation			
	Image		**Pixel**	
Méthodes \ **Critères**	**AUC**	**EER**	**AUC**	**RD**
MDT temporel[6]	0,82	0,23	0,57	0,59
MDT spatial[6]	0,6	0,43	0,66	0,5
Social Force[6]	0,69	0,36	0,22	0,41
MPPCA[6]	0,67	0,35	0,22	0,23
Adam (MHL)[6]	0,63	0,38	0,16	0,42
Energy-based [7]	0,7	0,35	0,49	0,67
Zhang et al. [12]	0,86	0,19	0,76	0,74
SA-MHOF [5]	0,86	0,19	0,63	0,68
Notre Méthode	**0,78**	**0,28**	**0,58**	**0,56**

Les tableaux 1 et 2 montrent les performances de notre méthode comparées à celles de la littérature. De ces résultats, on peut remarquer que notre méthode surpasse certains travaux antérieurs comme la version spatiale du "Mixture of Dynamic Texture" (MDT), l'approche de la force sociale, la méthode basée sur MPPCA, la méthode d'Adam [6] et la méthode basée sur Energy [7] sur les deux bases. Cependant, comparée aux méthodes de Zhang et al. [12], SA-MHOF [5], TCP [13] et Amraee et al. [14] notre méthode est moins performante.

Les principaux avantages de notre méthode sont sa simplicité et sa rapidité. Le descripteur proposé est très simple à comprendre, à implémenter et rapide à calculer. Il faut noter que notre descripteur est utilisé pour la sélection des blocs, ce qui offre un avantage en terme de complexité par rapport à d'autres approches qui emploient des algorithmes différents pour la sélection des régions et la construction du modèle. Malheureusement, il sera difficile de comparer le temps de calcul de notre méthode avec d'autres en raison de l'accès difficile à leur code.

1. http://www.svcl.ucsd.edu/projects/anomaly/dataset.html

TABLE 2 – Résultats d'évaluation sur la base Ped 2

Méthodes \ Critères	Niveau d'évaluation			
	Image		Pixel	
	AUC	EER	AUC	RD
MDT temporel [6]	0,76	0,28	0,52	0,57
MDT spatial [6]	0,75	0,29	0,66	0,63
Social Force [6]	0,7	0,35	0,22	0,28
MPPCA [6]	0,71	0,36	0,22	0,22
Adam (MHL) [6]	0,58	0,46	0,16	0,22
Energy-based [7]	0,86	0,16	0,72	0,84
TCP [13]	0,88	0,18	-	-
Amraee et al. [14]	0,93	-	0,85	-
Notre Méthode	**0,8**	**0,3**	**0,79**	**0,69**

4 Conclusion

Dans cet article, nous avons proposé une nouvelle méthode qui localise les événements rares dans une vidéo. L'approche proposée utilise la notion de signature de flot qui correspond à la fonction de signe de la transformée en cosinus discret d'une image. Une carte des mouvements saillants est générée via la transformée inverse du TCD de la signature des données du flot optique. Nous avons proposé d'utiliser cette dernière carte pour sélectionner les régions saillantes de la scène mais aussi de modéliser les événements. La modélisation des événements fréquents se fait avec l'algorithme SVM "one-class". Nous avons effectué deux niveaux d'évaluation basés sur le score de classification des régions. Les résultats expérimentaux montrent que la méthode proposée est compétitive par rapport aux travaux antérieurs. Elle offre une bonne localisation sur la base Ped 2 mais est mitigée pour sur la base Ped 1.

5 Remerciements

Ce travail fait partie du projet LUMINEUX, soutenu par la Région Centre-Val de Loire (France). Les auteurs tiennent à remercier le Conseil Régional du Centre - Val de Loire pour son soutien.

Références

[1] Vijay Mahadevan, Weixin Li, Viral Bhalodia, and Nuno Vasconcelos. Anomaly detection in crowded scenes. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1975–1981. IEEE, 2010.

[2] Mary W Green. The appropriate and effective use of security technologies in us schools : A guide for schools and law enforcement agencies series. *NCJ*, 1999.

[3] Myo Thida, Yoke Leng Yong, Pau Climent-Pérez, Howlung Eng, and Paolo Remagnino. A literature review on video analytics of crowded scenes. In *Intelligent Multimedia Surveillance*, pages 17–36. Springer, 2013.

[4] Yanjiao Shi, Yunxiang Liu, Qing Zhang, Yugen Yi, and Wenju Li. Saliency-based abnormal event detection in crowded scenes. *Journal of Electronic Imaging*, 25(6) :061608, 2016.

[5] Lei Hu and Fangyu Hu. Anomaly detection in crowded scenes via sa-mhof and sparse combination. In *Computational Intelligence and Design (ISCID), 2017 10th International Symposium on*, volume 1, pages 421–424. IEEE, 2017.

[6] Weixin Li, Vijay Mahadevan, and Nuno Vasconcelos. Anomaly detection and localization in crowded scenes. *IEEE transactions on pattern analysis and machine intelligence*, 36(1) :18–32, 2014.

[7] Hung Vu, Dinh Phung, Tu Dinh Nguyen, Anthony Trevors, and Svetha Venkatesh. Energy-based models for video anomaly detection. *arXiv preprint arXiv :1708.05211*, 2017.

[8] Yachuang Feng, Yuan Yuan, and Xiaoqiang Lu. Learning deep event models for crowd anomaly detection. *Neurocomputing*, 219 :548–556, 2017.

[9] Xiaodi Hou, Jonathan Harel, and Christof Koch. Image signature : Highlighting sparse salient regions. *IEEE transactions on pattern analysis and machine intelligence*, 34(1) :194–201, 2012.

[10] Berthold KP Horn and Brian G Schunck. Determining optical flow. *Artificial intelligence*, 17(1-3) :185–203, 1981.

[11] Xiaodi Hou and Liqing Zhang. Saliency detection : A spectral residual approach. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.

[12] Xinfeng Zhang, Su Yang, Xinjian Zhang, Weishan Zhang, and Jiulong Zhang. Anomaly detection and localization in crowded scenes by motion-field shape description and similarity-based statistical learning. *arXiv preprint arXiv :1805.10620*, 2018.

[13] Mahdyar Ravanbakhsh, Moin Nabi, Hossein Mousavi, Enver Sangineto, and Nicu Sebe. Plug-and-play cnn for crowd motion analysis : An application in abnormal event detection. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1689–1698. IEEE, 2018.

[14] Somaieh Amraee, Abbas Vafaei, Kamal Jamshidi, and Peyman Adibi. Abnormal event detection in crowded scenes using one-class svm. *Signal, Image and Video Processing*, pages 1–9, 2018.