



HAL
open science

Points d'intérêt spatio-temporels pour la détection de mouvements dans les vidéos

Alain Simac-Lejeune, Michèle Rombaut, Patrick Lambert

► To cite this version:

Alain Simac-Lejeune, Michèle Rombaut, Patrick Lambert. Points d'intérêt spatio-temporels pour la détection de mouvements dans les vidéos. MajecSTIC 2010 - MANifestation des JEunes Chercheurs en Sciences et Technologies de l'Information et de la Communication, Oct 2010, Bordeaux, France. pp.n.c. hal-00527382

HAL Id: hal-00527382

<https://hal.science/hal-00527382>

Submitted on 19 Oct 2010

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.

Points d'intérêt spatio-temporels pour la détection de mouvements dans les vidéos

Alain Simac-Lejeune^{1,2}, Michèle Rombaut² et Patrick Lambert¹

1 : Listic (Université de Savoie), BP 80439, 74944 Annecy-le-Vieux Cedex, France.

2 : Gipsa-lab (Université Grenoble 1), 961 rue de la Houille Blanche, BP 46, F-38402 Grenoble Cedex, France.

Contact : alain.simac-lejeune@univ-savoie.fr

Résumé

Parmi toutes les caractéristiques qui peuvent être extraites de vidéos, les points d'intérêt spatio-temporels (STIP) sont particulièrement intéressants car ce sont des caractéristiques de bas niveau simples et robustes qui permettent une bonne caractérisation des objets en mouvement. Dans cet article, nous définissons les STIP et analysons leurs propriétés. Puis, les STIP sont utilisés pour détecter des objets en mouvement et pour caractériser les changements spécifiques dans les mouvements de ces objets. Les performances sont étudiées sur des types très différents de vidéos : des séquences d'athlétisme et des séquences de films d'animation.

Abstract

Among all the features which can be extracted from videos, we propose to use Space-Time Interest Points (STIP). STIP are particularly interesting because they are simple and robust low-level features providing an efficient characterization of moving objects within videos. In this paper, after defining STIP and after giving some of their properties, we will use STIP to detect moving objects and to characterize specific changes in the movements of these objects. Proposed results are obtained from very different types of videos, namely athletic videos and animation movies.

Mots-clés : points d'intérêt, spatio-temporel, analyse de mouvement, vidéos.

Keywords: interest point, spatio-temporal, motion analysis, video.

1. INTRODUCTION

Le système de perception humaine est naturellement attiré par les mouvements et les déplacements des objets. Dans le cadre d'indexation vidéo, il semble plus pertinent d'analyser les zones de l'image dans lesquelles apparaît du mouvement, d'autant si celui-ci n'est pas constant. Les points d'intérêt spatio-temporels peuvent permettre de détecter ces zones d'intérêt et donc fournir de l'information utile quant au contenu sémantique de la vidéo. Différentes méthodes ont été proposées pour extraire des points d'intérêt spatiaux. Une évaluation de ces approches a été proposée dans [10]. Dans [9], Laptev et Lindeberg proposent une extension spatio-temporelle du détecteur de points d'intérêt proposé par Harris dans [5], intitulé "Space-Time Interest Points" et noté STIP dans la suite de l'article. Les STIP sont des points d'intérêt qui sont intéressants à la fois dans les domaines spatial et temporel. Les STIP ont été utilisés pour la reconnaissance d'actions [6], le résumé automatique [7] ou, plus généralement pour la détection d'événements spatio-temporels [8]. Dans cet article, nous proposons d'utiliser les STIP pour détecter des objets mobiles dans une scène et de caractériser des changements spécifiques dans le mouvement de ces objets. Pour illustrer la robustesse de cette approche, nous avons utilisé plusieurs types de vidéos très différents : des vidéos synthétiques, des vidéos de sport, des films d'animation... Dans la section 2, nous décrivons brièvement les vidéos qui sont utilisées dans cette étude. La section 3 présente les STIP ainsi que leurs propriétés. Les sections 4 et 5 montrent quelques résultats obtenus respectivement sur la détection des objets en mouvement et sur la localisation des changements dans des mouvements spécifiques.

2. BASE DE DONNÉES

Afin de caractériser les performances des approches proposées, nous avons utilisé quatre types différents de données vidéos :

- des vidéos de synthèse : 60 séquences composées d'images de synthèse d'une résolution de 288x288 pixels avec un fond uniforme noir et un ou plusieurs objets (ronds, carrés, triangles et formes composées de plusieurs lignes) en mouvement uniforme ou non, en mouvement rapide ou non.
- des vidéos de sports : 40 séquences de sauts d'athlétisme durant entre 100 et 160 images (environ 5 secondes) avec une résolution de 300x300 pixels. Ces séquences présentent des caractéristiques particulières : l'athlète est suivi par la caméra mobile et il reste généralement au centre de l'image. Le fond défile donc rapidement.
- un film d'animation¹, intitulé "Le Moine et le Poisson", qui dure 6 minutes et 23 secondes (5745 images) avec une résolution de 320x240 pixels.
- des vidéos hétérogènes issues de la télévision : dans ces séquences, on trouve notamment des séquences de hand-ball, de surveillance de circulation automobile. Ces séquences ont été utilisées dans [11].

Dans tous les tests présentés, les performances ont été évaluées sur des séquences préparées sans changement de plan, ceux-ci générant beaucoup de points d'intérêt. En effet, les tests effectués sur des vidéos ou sur des films montrent que les transitions peuvent être facilement détectées car elles se traduisent par un nombre important de points d'intérêt.

3. POINTS D'INTÉRÊT SPATIO-TEMPORELS

3.1. Présentation

Les points d'intérêt spatiaux (Spatial Interest Points - SIP) sont définis comme des points où apparaît un changement significatif dans l'image. Par exemple, les coins, les intersections, les points isolés et les points spécifiques sur les textures sont des points d'intérêt. En pratique, ces points d'intérêt correspondent à un pixel présentant un fort rayon de courbure de l'intensité c'est à dire des variations du second ordre de celle-ci.

Dans une séquence d'images, les points d'intérêt spatio-temporels (Space-Time Interest Points - STIP) sont des points présentant des changements significatifs dans le temps et dans l'espace. Ces points sont particulièrement intéressants car ils concentrent l'information initialement contenu dans toute l'image en quelques points spécifiques. L'intégration de la composante temporelle permet d'effectuer un filtrage sur les points d'intérêt spatiaux (SIP) et de ne conserver que ceux qui présentent également une discontinuité temporelle. Par exemple, le mouvement uniforme d'une voiture ne génère pas de points d'intérêt spatio-temporels tandis que les jambes d'une personne qui marche ou qui court en génèrent beaucoup à cause de la discontinuité des mouvements.

3.2. Points d'intérêt spatio-temporels : détecteur de Laptev

Laptev and al. [9] proposent une extension spatio-temporelle du détecteur de SIP de Harris [5] pour détecter les "Space-Time Interest Points", noté STIP dans la suite de l'article. La détection des STIP est réalisée en utilisant une matrice Hessienne-Laplace H [8] définie, pour un pixel (x, y) au temps t d'intensité $I(x, y, t)$, par :

$$H(x, y, t) = \begin{pmatrix} \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial x \partial y} & \frac{\partial^2 I}{\partial x \partial t} \\ \frac{\partial^2 I}{\partial x \partial y} & \frac{\partial^2 I}{\partial y^2} & \frac{\partial^2 I}{\partial y \partial t} \\ \frac{\partial^2 I}{\partial x \partial t} & \frac{\partial^2 I}{\partial y \partial t} & \frac{\partial^2 I}{\partial t^2} \end{pmatrix} \quad (1)$$

De manière analogue au détecteur de Harris, un filtre gaussien est appliqué dans le domaine spatial et dans le domaine temporel. Les deux paramètres σ_s et σ_t , contrôlent les échelles spatiale et temporelle pour les coins détectés. Pour extraire les STIP, différents critères ont été proposés. Comme dans [8], nous avons choisi d'utiliser l'extension spatio-temporelle de la fonction d'extraction R, appelée "saliency function", définie par :

$$R(x, y, t) = \det(H(x, y, t)) - k \times \text{trace}(H(x, y, t))^3 \quad (2)$$

1. Festival International du Film d'Animation d'Annecy FIFA - CITIA - <http://www.annecy.org/>

où le paramètre k est ajusté de manière empirique à 0.04 comme pour la détection des SIP. Les STIP correspondent aux plus grandes valeurs de la fonction d'extraction R en utilisant une valeur de seuil. Une valeur typique pour ce seuil est 150.

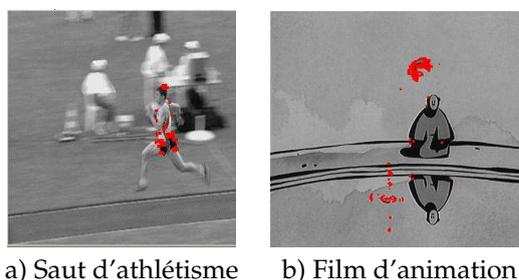
3.3. Implémentation

Pour des raisons d'échelle et de bruit induit par la fonction dérivée seconde, la dérivée (eq. 1) est calculée par convolution de l'image avec un filtre Gaussien approprié. Il est essentiel de séparer les filtres Gaussiens des composantes spatiale et temporelle qui sont généralement différentes et indépendantes. Comme dans [3], la composante spatiale n'est pas directement liée à la composante temporelle. Le terme $\frac{\partial^2 I}{\partial x \partial t}$ est obtenu par :

$$\frac{\partial g_{\sigma_s}(x, y)}{\partial x} \otimes \left(\frac{\partial g_{\sigma_t}(t)}{\partial t} \otimes I(x, y, t) \right) \quad (3)$$

où $g_{\sigma_s}(x, y)$ est le filtre Gaussien 2D de variance σ_s^2 et $g_{\sigma_t}(t)$ est le filtre Gaussien 1D de variance σ_t^2 ; \otimes est l'opérateur de convolution. Les autres termes de la matrice sont obtenus de manière similaire. La variance des filtres Gaussiens permet de contrôler les échelles spatiale temporelle des dérivées à évaluer.

Nous avons réalisé des tests de performance en analysant à la fois le nombre de STIP détectés et le taux de bonne détection. Une bonne détection correspond à un STIP détecté sur un objet réellement en mouvement. La figure 1 montre les STIP extraits sur deux exemples de séquences, l'une correspondant à un saut d'athlétisme et l'autre à un film d'animation. Le tableau 1 montre l'in-



a) Saut d'athlétisme b) Film d'animation

FIGURE 1 – Les STIP détectés sont représentés en rouge sur une image de la vidéo en noir et blanc. Sur l'image exemple (a) du saut d'athlétisme, les STIP sont situés sur l'athlète. Sur l'image exemple (b) du film d'animation, les STIP sont situés sur le poisson, sur l'eau (mouvement à la surface, gouttes) et sur le moine.

fluence des différentes valeurs d'écart type σ_s et σ_t . Ces résultats mettent en évidence l'impact des filtres Gaussien : quand les valeurs de σ_s et σ_t sont basses, le nombre de STIP augmentent, mais le taux de bonnes détections diminue. Au contraire, quand les valeurs de σ_s et σ_t sont hautes, le nombre de STIP décroît et le taux de bonnes détections augmente pour atteindre 100%. Néanmoins, les réglages correspondant au taux de 100% correspondent également aux plus faibles nombres de points générés. Au final, un bon compromis semble être $\sigma_s = 1.5$ et $\sigma_t = 1.5$. Bien qu'il existe des méthodes pour effectuer un réglage automatique (à partir d'une référence), nous avons préféré les définir manuellement afin d'optimiser les temps de calcul.

3.4. Propriétés

Les STIP possèdent des propriétés qui sont bien connues notamment leur stabilité relative par rapport aux transformations géométriques. Nous avons étudié d'autres propriétés qui sont aussi très importantes dans le cadre de l'analyse de vidéos, telles que la robustesse des STIP, au bruit impulsionnel, aux modifications de contraste, à l'orientation et au mouvement de caméra.

σ_s	σ_t	STIP/image	Precision
1.5	1.5	68	96,2%
1.2	1.5	213	84,4%
2.0	1.5	9	100%
1.5	1.2	191	85,7%
1.5	2.0	19	99,7%
1.2	1.2	521	71,9%
2.0	2.0	1	100%

Test - 20 séquences (8 sauts, 2 animations, 10 divers) - 2400 images
 $k = 0,04$ et $S = 150$

TABLE 1 – Influence des écarts types σ_s et σ_t sur les performances.

3.4.1. Orientation et mouvement de la caméra

Le tableau 2 présente les taux de détection obtenus pour différentes prises de vue. Ces tests ont été réalisés sur des séquences de saut d'athlétisme avec caméra mobile et prise de vue frontale ou latérale. Un STIP est considéré comme bien détecté s'il se trouve positionné sur l'athlète. Le contrôle a été effectué manuellement. Les résultats présentés dans le tableau 2 montrent que les

	STIP/frame	Bonne détection
Caméra de face	76	91%
Caméra de profil	91	99%
Caméra statique	82	93%
Caméra mobile	83	98%

Test - 20 séquences des différents sauts d'athlétisme (2360 images)
 $k = 0,04$, $\sigma_s = \sigma_t = 1.5$ et $S = 150$

TABLE 2 – Influence des propriétés de prises de vue sur le nombre de STIP

conditions de prise de vue ont une influence sur le nombre de STIP détectés.

Logiquement, la meilleure orientation est la prise latérale et un fond dynamique obtient une meilleure précision qu'un fond statique puisqu'il génère moins de fausses alarmes.

Mais, la surprise vient du test effectué sur le fond puisque le fond statique obtient des résultats légèrement moins bons que pour un fond défilant.

3.4.2. Variations de contraste et de bruit

Une analyse des modifications de qualité d'image sur la détection STIP a également été réalisée. Deux situations ont été examinées : des modifications de contraste et l'ajout de bruit de type impulsif, celui-ci étant le plus difficile à corriger. De plus, il modifie la détection de points d'intérêt. Le tableau 3(a) indique le nombre de STIP obtenus pour différentes valeurs de contraste et 3(b) le nombre de STIP obtenus pour différents niveaux de bruit. L'évaluation est réalisée en observant les variations du nombre de STIP par rapport à la situation initiale sans modification de contraste et sans bruit. On remarque que la détection STIP est très sensible à la modification du contraste. En revanche, le nombre de STIP est relativement stable par rapport au bruit impulsif (Figure 2) lorsque celui-ci n'est pas trop élevé.

3.4.3. Effet de la compression

Le dernier critère ayant une influence sur le nombre de STIP générés est le facteur de compression de la vidéo. En effet, la compression crée un effet d'aliasing, de crénelage des lignes, qui peuvent être perçues comme des angles [2]. Le tableau 4 présente l'influence du facteur de compression

Contraste	STIP/image	Bruit		STIP/image
		Puissance	Intensité	
25%	0	0	0	68
50%	< 1	20	20	69
75%	11	20	50	70
100% reference	68	20	70	72
125%	155	20	100	74
150%	203	50	20	83
175%	231	50	50	119
		50	70	153
		50	100	224

a) Influence du contraste

b) Influence du bruit

Test - 20 séquences de différents sauts d'athlétisme (2360 images)

 $k = 0,04$, $\sigma_s = \sigma_t = 1.5$ et $S = 150$

TABLE 3 – Influence du contraste et du bruit sur la qualité de détection des STIP.

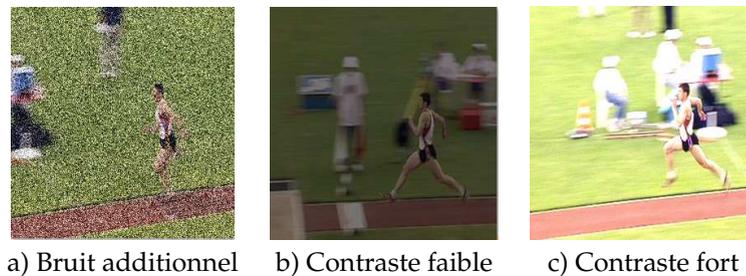


FIGURE 2 – Exemple de vidéos

MPEG2 sur le nombre de STIP générés. Ces tests ont été effectués sur des vidéos de synthèse et de sauts d'athlétisme. Il est important de noter que le carré ne génère pas de faux positif et donc l'aliasing n'a pas d'effet sur le nombre de STIP détectés. Ces résultats montrent que le facteur de compression a un effet important dès que l'on dépasse le seuil de 30% de compression. Afin de ne pas perturber les résultats, il sera donc nécessaire de s'assurer que les séquences utilisées ne sont pas compressées avec un facteur plus important que 30%.

Facteur	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
STIP/image	29	29	29	30	38	44	51	62	77	90	118

80 séquences de synthèse et de sauts d'athlétisme

 $k = 0,04$, $\sigma_s = \sigma_t = 1.5$ et $S = 120$

TABLE 4 – Influence du facteur de compression MPEG2

4. DÉTECTION D'OBJETS EN MOUVEMENT

Un objet est dit en mouvement dans une vidéo quand il se déplace par rapport au fond de l'image. Il existe de nombreuses méthodes pour détecter les objets en mouvement, basées sur la détection de mouvement [4], la segmentation [1], la différence entre images successives... Nous proposons

d'utiliser les STIP pour détecter ce type d'objets. Toutefois, cette méthode ne fonctionne que si l'objet a un mouvement non régulier car les STIP correspondent à la variation du second ordre dans l'espace et le temps. Tout d'abord, cette technique a été évaluée sur des vidéos de synthèse afin de pouvoir maîtriser les conditions expérimentales. La table de résultats 6 montre que les cercles ne génèrent aucun STIP, contrairement aux carrés et aux triangles qui en génèrent à leurs coins. Il est à noter que les vidéos composées de lignes brisées génèrent parfois des points en plus. Pour les autres types de vidéos (sauts d'athlétisme, films d'animation), les mouvements observés

	cercles	carrés	triangles	lignes brisées
Précision	N/A	1.00	1.00	0.99
Rappel	N/A	1.00	1.00	0.99

60 séquences de synthèse
 $k = 0,04, \sigma_s, \sigma_t = 1.5$ et $S = 120$

TABLE 5 – Performances de détection (images de synthèse)

	Précision	Rappel
Films d'animation	0.99	0.91
Vidéos d'athlétisme	0.99	0.95

20 séquences de saut en longueur (durée : 2120 images)
 et 500 images issues du film "Le Moine et le Poisson"
 $k = 0,04, \sigma_s, \sigma_t = 1.5$ et $S = 120$

TABLE 6 – Performances de détection d'objets en mouvement

concernent généralement un objet ou une personne qui a une rôle important dans la scène. Les tests sont effectués en utilisant le critère classique rappel/précision. La validation a été obtenue manuellement de la manière suivante :

- vrai positif : au moins un point sur un objet d'intérêt en mouvement,
- faux positif : au moins un point sur un objet qui n'est pas en mouvement,
- faux négatif : aucun point sur un objet d'intérêt en mouvement.

Le tableau 6 montre qu'on obtient de très bons résultats en utilisant les STIP comme détecteur d'objets intéressants, même dans les cas où plusieurs objets se déplacent dans la même image. Pour conclure, nous pouvons souligner que les STIP montrent des performances intéressantes pour détecter des objets en mouvement. Ils pourraient donc être utilisés lors d'une phase de pré-traitement suivie par une méthode de localisation, et éventuellement de segmentation et de suivi.

5. DÉTECTION DE CHANGEMENTS DANS LES MOUVEMENTS

5.1. Principe

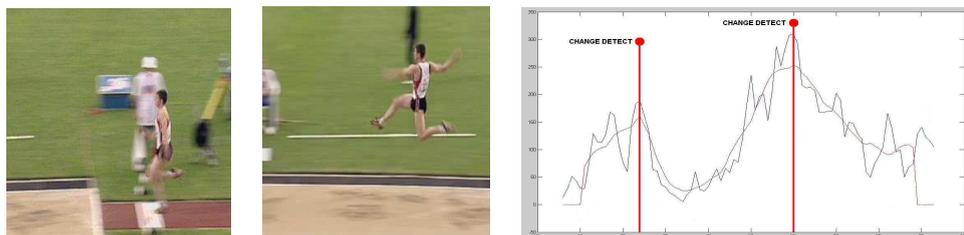
Dans [7] Laganier définit le niveau d'activité au sein d'une vidéo comme le nombre de pixels qui sont modifiés entre deux images. Il propose donc de définir une fonction "activité" comme étant le nombre de STIP détectés par image. L'évolution temporelle de cette activité peut contenir des informations intéressantes d'un point de vue sémantique. En particulier, les maxima locaux de cette fonction d'activité sont généralement liés à des événements importants dans la séquence tels qu'une transition entre deux phases d'un même mouvement. Par exemple, un saut d'athlétisme peut être décomposé en une phase de course à pied puis, après un appel, la phase ascendante suivie de la phase descendante du saut.

La fonction d'activité est généralement très bruitée, elle est donc d'abord lissée grâce à l'utilisation d'un filtre moyenneur de taille de $\alpha * 2 + 1$ (pour un filtre avec un $\sigma_t = 1,5$, on prend 11 images donc $\alpha = 5$) correspond au nombre d'images nécessaire pour le calcul des STIP d'une image. On note $a_{filt}(t)$ la fonction d'activité filtrée. Puis on effectue la détection des maxima locaux de

$a_{filt}(t)$ qui satisfont la condition suivante :

$$0.8 \times a_{filt}(t - \alpha) \leq a_{filt}(t) \leq 0.8 \times a_{filt}(t + \alpha) \quad (4)$$

La figure 3 illustre la détection de deux transitions au cours d'un saut en longueur : la première est l'appel du saut et la seconde est le moment où l'athlète est au plus haut du saut.



La courbe bleue est le niveau d'activité et la rouge, le niveau d'activité lissé

FIGURE 3 – Détection des transitions lors d'un saut en longueur

5.2. Résultats expérimentaux

Nous avons tout d'abord utilisé 60 séquences de synthèse pour effectuer la qualification de la méthode. L'évaluation porte sur la détection des changements d'angle de 90° ou des changements de sens dans les trajectoires des objets. Le tableau 7 présente les résultats. Globalement, les transitions sont bien détectées avec une précision temporelle de 3 images environ. Nous avons ensuite

	Précision	Rappel	Tolérance
90°	1.00	1.00	± 3 frames
Inversion	1.00	1.00	± 3 frames
Tous	1.00	0.98	± 3 frames

60 séquences de synthèse

$k = 0,04$, $\sigma_s = \sigma_t = 1.5$ et $S = 120$

TABLE 7 – Performances de détection de changements significatifs dans les séquences de synthèse

utilisé une vingtaine de séquences de différents types de sauts (saut en hauteur, saut à la perche, saut en longueur et triple saut), une dizaine de séquence de hand-ball et une dizaine de séquence de surveillance routière. Pour les sauts d'athlétisme et pour la surveillance routière, les séquences contiennent généralement un seul événement. L'évaluation est une comparaison entre les transitions de la vérité terrain et celles détectées. Comme l'emplacement temporelle d'une transition est n'est pas toujours obtenue de manière précise, nous avons accepté une tolérance sur leurs positions. Sur le tableau 8, on peut observer que les transitions sont correctement détectées avec une précision comprise entre 3 et 10 images. Les valeurs de précision et de rappel sont relativement élevées. Les performances sont moins satisfaisantes pour le triple saut, le changement de mouvement étant moins marqué. Concernant les vidéos de hand-ball, les changements correspondant aux passes entre les joueurs sont en général visibles.

6. CONCLUSIONS

La détection de STIP montre des résultats probants pour la détection d'objets en mouvement et pour la détection de phases dans les mouvements. Toutefois, elle présente quelques limitations.

	Précision	Rappel	Tolérance
Saut en longueur	0.93	0.92	± 3 frames
Saut en hauteur	0.92	0.88	± 3 frames
Triple saut	0.81	0.71	± 5 frames
Saut à la perche	0.84	0.85	± 10 frames
Hand-ball	0.76	0.86	± 3 frames
Circulation automobile	0.97	0.93	± 3 frames

20 séquences de saut, 8 séquences de hand-ball et 8 de surveillance routière
 $k = 0,04$, $\sigma_s = \sigma_t = 1.5$ et $S = 120$

TABLE 8 – Performances de détection de changements significatifs dans les mouvements

La première concerne le paramétrage. En effet, les paramètres σ_s et σ_t sont difficiles à régler et très dépendants du type de vidéo et de leurs caractéristiques. Dans la poursuite de travaux, il sera intéressant de trouver une méthode d'adaptation automatique des paramètres afin de garantir la stabilité des points générés. La deuxième limite repose sur les conditions de prise de vue et la qualité vidéo (bruit, contraste et compression), en particulier dans le cas des vidéos à diffusion temps réel. Ces contraintes peuvent être problématiques si l'on veut utiliser cet outil pour des vidéos sur le Web ou des vidéos en streaming. Dans ce cas, il sera probablement nécessaire de faire un pré-traitement pour ajuster le contraste et / ou de filtrer le bruit. La dernière limite porte sur la fiabilité. Dans le cas d'un mouvement de vitesse faible ou constant (pour la détection d'objet) ou dans le cas d'un mouvement dont les changements ne sont pas assez significatifs pour détecter les phases du mouvement, les performances seront dégradées. Cependant, la force de cette approche est qu'elle permet d'obtenir des résultats en un temps de calcul très faible ce qui permet de l'utiliser comme une phase de pré-traitement pour focaliser les algorithmes de traitement sur des zones d'intérêt.

Bibliographie

1. A. Bugeau. *Détection et suivi d'objets en mouvement dans des scènes complexes, application à la surveillance des conducteurs*. Thèse de doctorat, IRISA, 2007.
2. R.J. Clarke. *Digital compression of still images and video*. London : Academic press, pages 285–299, 1995.
3. P. Dollar, V. Rabaud, G. Cottrell, et S. Belongie. Behaviour recognition via sparse spatio-temporal interest point detector. *VSPETS*, 2005.
4. B. Gai-Checa, P. Boutheymy, et T. Vieville. *Détection d'objets en mouvement*. Technical Report INRIA-RR - 1906, INRIA, 1993.
5. C. Harris et M.J. Stephens. A combined corner and edge detector. *In Alvey Vision Conference*, 1988.
6. Y. Ke, R. Sukthankar, et M. Hebert. Efficient visual event detection using volumetric features. *In International Conference on Computer Vision*, pages 166–173, 2005.
7. R. Laganière, R. Bacco, A. Hocevar, P. Lambert, G. Païs, et B.E. Ionescu. Video summarization from spatio-temporal features. *ACM*, 2008.
8. I. Laptev. On space-time interest points. *International Journal of Computer Vision*, 64(2/3) :107–123, 2005.
9. I. Laptev et T. Lindeberg. Space-time interest points. *ICCV'03*, pages 432–439, 2003.
10. C. Schmid, R. Mohr, et C. Bauckhage. Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(2) :151–172, 2000.
11. A. Simac-Lejeune, S. Marat, D. Pellerin, P. Lambert, M. Rombat, et N. Guyader. Relevance of interest points for eye position prediction on videos. *ICVS*, 2009.