

Approche d'Evaluation des Techniques de Suivi d'Objet

A. GANOUN^{1,2} R. CANALS¹ R. LECONGE¹

¹Laboratoire d'Electronique, Signaux, Images - Université d'Orléans - 12 rue de Blois - 45067 ORLÉANS

²Faculté d'Ingénierie, Université de Garyounis, Benghazi, LIBYE

{ ali.ganoun ; raphael.canals ; remy.leconge } @univ-orleans.fr

Résumé – Dans cet article, nous nous intéressons à la manière d'évaluer les performances des techniques de suivi, ainsi que de savoir comment obtenir un suivi robuste et de qualité. Une approche d'évaluation est appliquée sur trois techniques de suivi basées sur l'approche de CamShift. Les résultats de chaque technique de suivi sur plusieurs séquences d'images sont évalués et comparés en termes d'erreur de suivi, de robustesse du suivi et de temps de calcul.

Abstract – In this article, we are interested in the method of evaluating the performances of tracking techniques and how to get an accurate and robust tracking. An evaluation technique is applied on three tracking techniques based on the CamShift approach. The results of each technique on several image sequences are evaluated and compared in terms of tracking error, tracking robustness and computational cost.

1. INTRODUCTION

Dans la littérature, de nombreuses méthodes ont été développées pour résoudre le problème du suivi d'objet, s'étendant d'approches simples jusqu'aux plus complexes [1], [2]. Une fois les différentes solutions définies, l'évaluation des performances des algorithmes de suivi est une étape importante pour comparer et améliorer les techniques de suivi, [3], [4], [5], [6], [7], [8]. Dans [6], les auteurs proposent une approche d'évaluation des performances pour un système d'assistance à la conduite, ceci en quantifiant la précision de position du véhicule sur la route. Un certain nombre de techniques d'évaluation de performance fonctionnent sans avoir à réaliser de comparatif avec la vérité terrain (Ground Truth GT) [9]. Dans [3], les auteurs proposent une évaluation de performance quantitative de techniques de segmentation et de suivi en employant des différences spatiales de couleur et de mouvement le long des contours des objets. La couleur et les mesures de mouvement sont combinées pour obtenir des notes qui reflètent le succès de la segmentation et du suivi. Pour valider leur technique, ils proposent une analyse de corrélation canonique entre leurs résultats de performance et des mesures obtenues avec le GT.

Nous proposons ici d'évaluer les performances des techniques de suivi de la même manière que dans [5], en considérant les mesures suivantes : l'erreur de suivi, la robustesse du suivi et le temps de traitement. Nous appliquons notre approche sur des techniques de suivi basées sur l'approche de CamShift [10], [11]. Trois variantes du CamShift sont considérées selon la dimension de l'histogramme manipulé : un histogramme 3D [12] pour lequel les trois canaux de couleur sont utilisés ; un histogramme 2D [13] où seuls 2 canaux sont utilisés : le plus représentatif de la cible et le moins représentatif du fond ; un histogramme 1D [14] dans le contexte des

images en niveaux de gris. Dans la suite, la notation «CamShift 2D» se réfère à la technique de CamShift avec l'histogramme 2D.

Cet article est organisé comme suit : dans la section suivante, nous présentons les principes de notre approche d'évaluation : l'efficacité de calcul, les séquences de test, l'initialisation et les paramètres de configuration algorithmiques. Les résultats des comparaisons sont présentés en section 3, suivis d'une conclusion en section 4.

2. PRINCIPES D'EVALUATION

2.1 Performances du suivi

Afin de comparer l'efficacité de calcul des techniques de suivi, les algorithmes ont été implantés et testés. Pour évaluer la stabilité du suivi, chaque algorithme est répété 10 fois avec différentes zones d'initialisation. La variation des résultats est étudiée et analysée statistiquement ; les performances de chaque algorithme de suivi sont estimées en calculant l'erreur de suivi ϵ , le pourcentage de convergence ρ et le temps de traitement. L'erreur de suivi représente la déviation du résultat de l'algorithme par rapport au résultat optimal (GT défini manuellement), alors que le pourcentage de convergence correspond au prorata d'essais pour lesquels l'algorithme fournit un résultat correct. Les temps d'exécution pris pour chaque essai sont calculés et moyennés.

2.2 Séquences de test

De nombreux tests ont été réalisés sur plusieurs séquences vidéo. Seules cinq séquences particulières sont présentées ici. La première séquence est la séquence Road dans laquelle l'objectif est de suivre un véhicule. La caractéristique principale de cette séquence se rapporte aux changements de taille de la cible au fil du temps. La

Tab.1 : Informations générales sur les séquences de test.

Nom de la séquence	Taille de l'image (pixels)	Nbre d'images	Type d'objet	Taille de la cible (pixels)	Propriétés de la séquence				
					Changement d'éclairage	Variation d'apparence	Occultation	Fusion et division	Changement de taille
Road	320x240	90	Véhicule	350	✓	✓	✗	✗	✓
HC1	720 x 240	244	Hélicoptère	4200	✓	✓	✗	✗	✓
OC1	720 x 576	175	Piéton	6500	✗	✓	✓	✗	✗
OC2	720 x 576	97	Piéton	2700	✗	✓	✓	✗	✗
Univ	720 x 576	114	Piéton	5000	✓	✓	✓	✗	✗

deuxième séquence est la séquence HC1 ; la cible est un hélicoptère dont l'apparence change : l'hélicoptère est souvent partiellement caché par ses propres pales.

Les trois autres séquences (OC1, OC2 et Univ) concernent le suivi de personnes. Dans la séquence OC1, l'apparence de la cible ne change pas beaucoup, mais celle-ci est partiellement occultée pendant environ 10 images. La séquence OC2 contient des occultations avec un cycliste, un piéton et deux voitures. Enfin la séquence Univ est une séquence de mauvaise qualité ; de plus, la cible subit de grandes occultations. Le Tableau 1 résume les informations générales de chacune de ces séquences.

2.3 Initialisation

L'initialisation, c.-à-d. la définition de l'objet à suivre dans la première image, joue un rôle important dans les algorithmes de suivi. Ainsi, pour considérer les variations de la sélection de la cible réalisée par un opérateur, une fenêtre d'initialisation est désignée de manière aléatoire autour de la cible, autant en taille qu'en position. Une détection de contours au moyen du filtre de Canny est lancée dans cette fenêtre, puis une fermeture des frontières externes, ce qui nous fournit le modèle de la cible à suivre dans la suite de la séquence. Pour la même séquence, chaque algorithme est répété 10 fois avec une fenêtre d'initialisation différente.

2.4 Paramètres de Configuration

Afin de comparer les résultats des différentes techniques de suivi, il est important de considérer l'effet des différentes combinaisons de paramètres de configuration pour chaque technique. Dans cette section, l'influence des différents paramètres pour toutes les séquences de test est évaluée ; l'objectif est de déterminer les paramètres qui fournissent les meilleures performances de suivi pour chaque séquence avec la technique de suivi correspondante.

Ainsi, pour chaque approche, en utilisant une séquence de test et une combinaison différente de paramètres, dix expériences indépendantes sont exécutées avec une initialisation différente. Les erreurs de suivi résultantes de ces dix tests sont évaluées de manière statistique. Cette analyse inclut les valeurs d'erreur minimale, maximale et moyenne. Elle inclut également l'analyse de variance ANOVA, une méthode d'analyse statistique indiquant si les différences des moyennes entre les groupes de données sont significatives. Si la sortie de cette méthode, la p-

valeur, est faible, les groupes de données sont alors différents. Les meilleurs paramètres sont finalement choisis suivant l'ordre d'importance décroissant suivant :

1. Le pourcentage de convergence le plus élevé ;
2. La valeur minimale de l'erreur moyenne de suivi ;
3. La valeur maximale de la p-valeur ;
4. Le coût en temps de calcul minimal.

Pour les techniques basées sur l'algorithme de Cam-Shift, trois paramètres principaux affectent l'erreur de suivi et le pourcentage de convergence. Le premier est le seuil ω à partir duquel nous appliquons la mise à jour du modèle, d'une image à l'autre. Ici, l'image de probabilité est une image en niveaux de gris. Quand la probabilité d'un pixel d'appartenir à la cible augmente, la valeur correspondante de l'image de probabilité croît : la valeur moyenne de l'image de probabilité monte donc quand la probabilité de trouver la cible est plus importante. Ce seuil permet de contrôler la valeur de la probabilité moyenne de présence de l'objet dans l'image à partir de laquelle la mise à jour du modèle est effectuée. Les seuils 0.1, 0.5 et 0.9 de ω sont considérés, chaque valeur représentant le rapport entre la moyenne calculée de la nouvelle image de probabilité et la moyenne de l'image de probabilité du modèle de la cible. Quand ω vaut 0.1, cela signifie que nous appliquons la mise à jour du modèle de la cible environ à chaque image. Cette situation est appropriée aux cas où il n'y a aucun changement de la cible pendant la séquence. Par contre, si le seuil augmente, la mise à jour du modèle de la cible est appliquée seulement dans les cas où la nouvelle image de probabilité est similaire à l'image du modèle.

Le deuxième paramètre est la plage de variation β de la taille de l'objet par rapport à la taille de l'objet dans l'image précédente. Nous considérons ici les valeurs 0.1, 0.5 et 0.9. Quand cette valeur est faible, la plage de variation admise est petite, ce qui convient quand la taille de l'objet ne change pas beaucoup. Quand cette valeur passe à 0.9, la plage de variation admise est alors très grande, ce qui est adapté à des changements considérables de la taille de l'objet entre deux images successives.

Pour prendre en compte l'évolution de la cible au fil du temps, il faut mettre à jour le modèle d'histogramme correspondant à notre modèle de cible. Dans le but d'améliorer cette phase de mise à jour, nous considérons la relation suivante :

$$H = \alpha H_p + (1 - \alpha) H_c \quad (1)$$

où H est le nouveau modèle d'histogramme, H_p est le modèle d'histogramme précédent, H_c est le modèle d'histo-

gramme dans l'image courante et α est le troisième paramètre indiquant le pourcentage du modèle d'histogramme précédent utilisé pour générer le nouveau modèle d'histogramme. Les valeurs suivantes sont traitées : 0.1, 0.5 et 0.9. Le cas $\alpha=0.1$ convient quand il n'y a aucune occultation et l'histogramme de l'objet à suivre change largement d'une image à la suivante.

Pour toutes les combinaisons de paramètres, les tests sont répétés 10 fois : 270 expériences sont donc réalisées sur chaque séquence.

3. RESULTATS

Pour illustration, les résultats de l'algorithme CamShift 2D avec la séquence de test OC1 sont donnés dans le Tableau 2. Ce tableau présente, pour quelques combinaisons de paramètres, l'analyse statistique de l'erreur de suivi (minimale, maximale, moyenne et la p-valeur), le pourcentage de convergence et le temps de calcul.

Deux exemples de courbes de performance sont exposés à la Figure 1, indiquant l'erreur de suivi à chaque image pour deux combinaisons des trois paramètres, ceci pour 10 essais (couleur). Elles permettent aussi de présenter la performance de suivi avec différentes valeurs de α et des valeurs fixées de β et ω pour un test spécifique (Figure 2).

Une autre représentation des résultats peut être faite avec des diagrammes en boîte à moustaches (Figure 3), chaque boîte correspondant à un test. Les diagrammes en boîte fournissent une visualisation concise de la distribution de diverses expériences, avec différentes combinaisons de paramètres. La ligne centrale dans chaque boîte marque la valeur moyenne et les extrémités de la boîte correspondent à l'écart-type ; les moustaches, ici, mènent aux valeurs minimales et maximales.

Dans la figure 4, deux exemples d'échec du CamShift avec la séquence Road avec différents paramètres sont présentés.

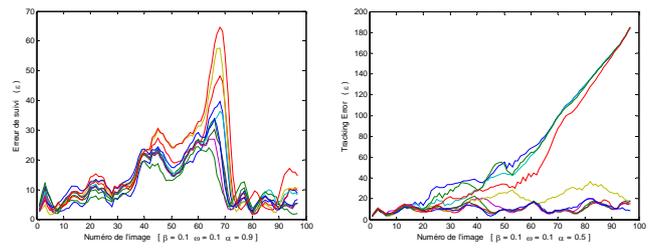
Pour la séquence Road, les trois techniques échouent dans le suivi de la cible, avec une p-valeur généralement faible. La raison principale de cet échec est la taille de la cible qui diminue au fur et à mesure que le véhicule s'éloigne : la petite taille de l'objet rend difficile la distinction entre la cible et le fond.

Bien que le mouvement de l'objet soit complexe et les changements d'éclairage de la scène considérables dans la séquence HC1, les 3 techniques permettent de suivre correctement la cible, avec des pourcentages de convergence atteignant 100. La raison de ce succès est la simplicité de la séquence avec une large différence de niveaux de gris entre la cible et le fond.

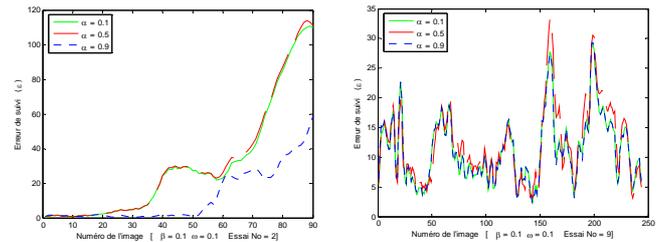
Dans la séquence OC1, les CamShift 3D et 2D sont les seules techniques qui permettent de suivre avec succès la cible. Ceci est dû en grande partie à la richesse de l'information traitée puisque ces deux techniques travaillent sur des images couleur : le pourcentage de convergence atteint 100 et la p-valeur est nulle, ce qui indique que chacune des deux techniques fournit les mêmes résultats sur les dix essais bien que les initialisations soient différentes.

Tab.2 : Résultats du CamShift 2D avec la séquence OC2 pour différentes combinaisons de paramètres.

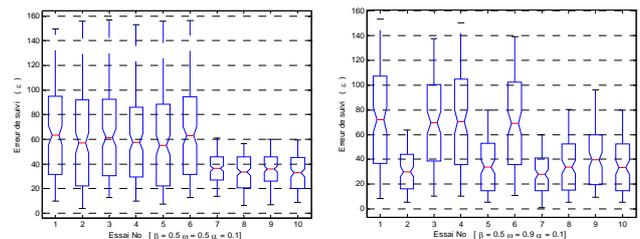
		$\alpha = 0.1$				$\alpha = 0.5$				$\alpha = 0.9$			
	ε	min	max	mean	p	min	max	mean	p	min	max	mean	p
		$\beta=0.1$	0.15	65	14	0	0.87	66	14.1	0	0.26	66	13.4
	p	100				100				100			
	t	1.41				0.43				1.4			
	ε	min	max	mean	p	min	max	mean	p	min	max	mean	p
		$\beta=0.1$	7.1	218	104	0	5.3	218	106	0.004	6.23	221	106
	p	0				0				0			
	t	0.39				1.36				1.38			
	ε	min	max	mean	p	min	max	mean	p	min	max	mean	p
		$\beta=0.5$	1.33	121	29	0	1.02	105	22	0	1	104	23.9
	p	100				100				100			
	t	1.35				1.36				1.36			
	ε	min	max	mean	p	min	max	mean	p	min	max	mean	p
		$\beta=0.5$	1.01	216	91	0	0.52	211	92	0	1.67	211	82
	p	10				10				20			
	t	1.48				1.48				1.47			
	ε	min	max	mean	p	min	max	mean	p	min	max	mean	p
		$\beta=0.9$	0.49	94.4	18.3	0	0.29	77.9	14	0	0.24	68.4	13.9
	p	60				80				70			
	t	1.44				1.42				1.42			



(CamShift 2D avec OC2) (CamShift 1D avec OC2)
Fig.1 : Exemples de courbes de performance



(CamShift 3D avec Road) (CamShift 2D avec HC1)
Fig.2 : Exemples de courbes de performance



(OC1 CamShift 3D) (Univ CamShift 2D)
Fig.3 : Exemples de diagrammes en boîte (ANOVA)

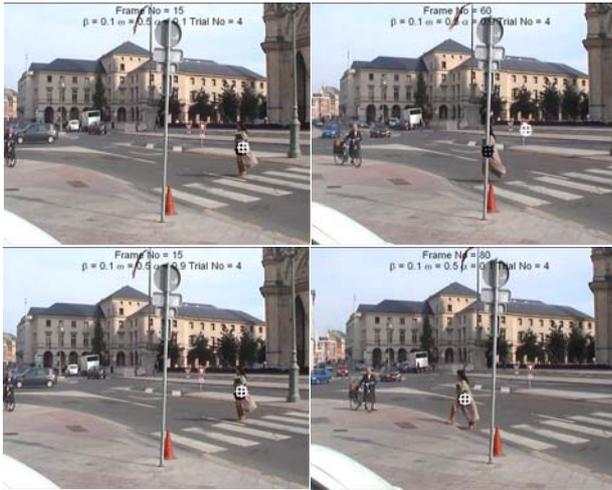


Fig.4 : Exemples d'échec (1ère ligne) et de succès du suivi (2ème ligne) de l'algorithme CamShift 2D avec OC2 (GT en noir, résultat du suivi en blanc).

Dans la séquence OC2, les CamShift 3D et 2D arrivent à suivre la cible avec un pourcentage de convergence de 100, et un petit pourcentage de 20 pour le CamShift 1D (le bon déroulement du suivi dépend de l'initialisation du processus). Il est clair que cette séquence est bien plus simple que les séquences OC1 et Univ au niveau de la complexité de la cible et du fond ainsi que de la manière dont se déroule l'occultation.

Tab.3 : Résultats optimaux de suivi avec les 3 techniques.

	CamShift 3D				CamShift 2D				CamShift 1D			
	min	max	mean	p	min	max	mean	p	min	max	mean	p
Road	0.06	117	28	0	0.14	78.2	16.3	0.06	1.75	34.8	18.1	0.29
	$\rho = 0$				$\rho = 0$				$\rho = 0$			
	$t = 0.28$				$t = 1.22$				$t = 0.22$			
	$\beta = 0.1 \ \omega = 0.1 \ \alpha = 0.1$				$\beta = 0.1 \ \omega = 0.1 \ \alpha = 0.1$				$\beta = 0.1 \ \omega = 0.1 \ \alpha = 0.5$			
HCI	0.47	45	12	0.99	0.42	93	34	0.85	0.35	33	9	0.98
	$\rho = 100$				$\rho = 100$				$\rho = 100$			
	$t = 0.37$				$t = 0.36$				$t = 0.24$			
	$\beta = 0.5 \ \omega = 0.1 \ \alpha = 0.5$				$\beta = 0.5 \ \omega = 0.9 \ \alpha = 0.1$				$\beta = 0.5 \ \omega = 0.1 \ \alpha = 0.9$			
OC1	0.13	61	22	0	0.45	77	15	0	0.95	325	43	0
	$\rho = 100$				$\rho = 100$				$\rho = 0$			
	$t = 0.41$				$t = 0.39$				$t = 1.05$			
	$\beta = 0.5 \ \omega = 0.1 \ \alpha = 0.5$				$\beta = 0.9 \ \omega = 0.1 \ \alpha = 0.5$				$\beta = 0.9 \ \omega = 0.5 \ \alpha = 0.1$			
OC2	2.5	65	19	0.02	0.24	47	13	0.24	0.67	790	20	0
	$\rho = 100$				$\rho = 100$				$\rho = 20$			
	$t = 1.41$				$t = 1.4$				$t = 0.42$			
	$\beta = 0.5 \ \omega = 0.1 \ \alpha = 0.9$				$\beta = 0.5 \ \omega = 0.1 \ \alpha = 0.9$				$\beta = 0.1 \ \omega = 0.1 \ \alpha = 0.9$			
Univ	1.7	101	38	0	1.0	153	48	0	3.44	120	55	0
	$\rho = 0$				$\rho = 20$				$\rho = 0$			
	$t = 1.2$				$t = 1.36$				$t = 0.72$			
	$\beta = 0.1 \ \omega = 0.5 \ \alpha = 0.1$				$\beta = 0.5 \ \omega = 0.9 \ \alpha = 0.1$				$\beta = 0.1 \ \omega = 0.9 \ \alpha = 0.9$			

La complexité de la séquence Univ explique l'échec des techniques dans le suivi de la cible. En fait, seul le CamShift 2D permet d'obtenir de bons résultats de suivi de la cible avec des pourcentages de convergence néanmoins assez faibles et une p-valeur nulle pour tous les cas.

Le Tableau 3 résume les paramètres et les résultats optimaux obtenus avec les trois techniques de suivi et les cinq séquences de test.

4. CONCLUSION

Dans cet article, nous avons présenté un cadre simple mais efficace pour évaluer les performances de trois techniques de suivi, évaluation basée sur l'étude directe et statistique de la déviation des résultats par rapport au GT. L'ensemble des données utilisées dans cette étude permet de considérer les différents éléments que nous attendons d'un algorithme de suivi d'objet. Concernant la comparaison des techniques de suivi d'objet dans des séquences avec un fond et/ou une cible complexes, le CamShift 2D génère les meilleurs résultats de suivi parmi les techniques étudiées.

Références

1. Yilmaz, A., O. Javed, and M. Shah, *Object Tracking: A Survey*, ACM Computing Surveys, 2006. **38**(4).
2. Nguyen, V. and Y. Tan, *Fast Block-Based Motion Estimation Using Integral Frames*. IEEE Signal Processing Letters, 2004. **11**(9).
3. Erdem, C., B. Sankur, and A. Tekalp, *Performance measures for video object segmentation and tracking*. IEEE Transactions on Image Processing, 2004. **13**(7).
4. Black, J., T. Ellis, and P. Rosin. *A Novel Method for Video Tracking Performance Evaluation*. in IEEE PETS, 2003, France.
5. Cavallaro, A. and F. Ziliani. *Characterisation of tracking performance*. in *6th International Workshop on Image Analysis for Multimedia Interactive Services*. 2005, Switzerland.
6. McCall, J.C. and M.M. Trivedi. *Performance evaluation of a vision based lane tracker designed for driver assistance systems*. in *IEEE Intelligent Vehicles Symposium*, 2005.
7. Pingali, S. and J. Segen. *Performance evaluation of people tracking systems*. in *IEEE WACV*, 1996.
8. Brown, L., et al. *Performance Evaluation of Surveillance Systems Under Varying Conditions*. in *IEEE PETS* 2005.
9. Correia, P.L. and F. Pereira, *Objective evaluation of video segmentation quality*. IEEE Transactions on Image Processing, 2003. **12**(2).
10. Intel Corporation, *Open Computer Vision Library, Reference Manual*. 2001.
11. Bradski, G., *Computer Vision Face Tracking For Use in a Perceptual User Interface*. Intel Technology Journal, 1998. **Q2**.
12. Allen, J., R. Xu, and J. Jin. *Object Tracking Using CamShift Algorithm and Multiple Quantized Feature Spaces*. in *Pan-Sydney Area Workshop on Visual Information Processing*. 2003, Australia.
13. Ould-Driss, N., A. Ganoun, and R. Canals. *Improved Object Tracking With Camshift Algorithm*. in *IEEE ICASSP*. 2006, France.
14. Ganoun, A., N. Ould-Driss, and R. Canals. *Tracking System Using CamShift and Feature Points*. in *EUSIPCO*, 2006, Italy.