Segmentation d'images infrarouge pour l'assistance aux personnes à domicile

Félix POLLA[1], Kamal BOUDJELABA[1], Bruno EMILE[1], Hélène LAURENT[2]
[1] Université d'Orléans, Laboratoire Prisme, Orléans, France
[2]INSA Centre Val de Loire, Laboratoire Prisme, Bourges, France
felix.polla1@univ-orleans.fr

Résumé

Dans cet article, nous montrons comment utiliser des informations spatiales d'une image provenant d'un capteur infrarouge, pour obtenir des résultats de segmentions permettant de localiser un objet dans une image. La méthode proposée exploite les résultats de segmentation de méthodes existantes pour améliorer la détection de la forme et du contour de l'objet en mouvement dans l'image. Dans une étude comparative, nous utilisons cinq méthodes de segmentations de l'état de l'art auxquelles nous ajoutons notre proposition afin d'obtenir de bonnes performances. Pour les images très spécifiques obtenues en sortie du capteur infrarouge, il ressort que l'utilisation de l'algorithme de segmentation par seuillage combiné à notre approche pour mieux caractériser un pixel comme arrière-plan ou non, aboutit au meilleur résultat. L'objectif visé à terme par ces travaux est l'exploitation de ces résultats dans un but de classification des actions des personnes présentes dans la scène.

Mots-clés: analyse spatio-temporelle, segmentation, évaluation.

I. INTRODUCTION

Contexte

Dans un contexte de vieillissement de la population, les solutions d'assistance aux personnes âgées (principalement à leur domicile) sont en plein essor. Des approches algorithmiques innovantes basées sur la vision par ordinateur ont été proposées ces dernières années. Le principal écueil auquel se heurtent les techniques développées dans le domaine visible, et ce malgré des performances intéressantes, est une réticence marquée des utilisateurs concernés à être filmés. Même si les vidéos ne sont pas enregistrées et jouent le rôle de capteurs fournissant des informations ciblées sur le comportement général (présence /absence, degré d'activité/ immobilité...), la peur d'une utilisation à des fins de surveillance et d'intimité dévoilée reste très présente**.** L'une des solutions envisagée par les acteurs des métiers de la domotique est l'utilisation de capteurs spécifiques.

Présentation du capteur

Le capteur développé par Irlynx[1] est un dispositif capable de retourner des images des objets chauds en mouvement dans une pièce. Ce capteur a la particularité de détecter un intrus dans l'obscurité la plus totale et aussi celle de ne pas permettre une reconnaissance et une identification individualisée des acteurs se trouvant dans la scène. La technologie développée repose sur le principe de la détection pyro-électrique, c'est-à-dire la détection du déplacement d'un corps chaud présent dans le volume surveillé. Ce capteur nous permet de travailler sur des images en vue de dessus, de taille 64×64 pixels.

Objectifs

L'objectif est d'analyser les comportements des personnes en utilisant les images provenant du capteur. Dans le processus d'analyse du comportement humain tel que présenté dans [1], la segmentation est l'étape qui vient après l'acquisition des données. Elle joue un rôle important dans le processus de décision car elle influe sur les autres étapes à savoir la classification, le suivi d'objet et la reconnaissance d'action ou de posture. La segmentation en traitement d'image consiste à partitionner l'image en différentes régions selon les objets qui la constituent (personnes, véhicules…). Nous nous intéressons ici à une segmentation binaire utilisant des informations spatiales de l'image. La segmentation binaire consiste à subdiviser l'image en deux régions à savoir la classe des pixels faisant partir de l'objet et ceux faisant partir de l'arrière-plan.

Plusieurs approches de segmentation d'objets en mouvement ont été proposées. Elles sont généralement regroupées en trois grandes catégories. L'approche de différence inter frame : cette approche détecte l'objet en faisant une différence entre deux ou trois frames consécutifs. Les pixels résultant de la soustraction doivent prendre des valeurs supérieures à un seuil pour être comptabilisés comme pixels mouvement par rapport aux pixels de l'arrière-plan [2,3]. Le principal inconvénient de ces approches est la perte de détection lorsqu'il n'y a plus de mouvement dans la scène observée.

L'approche de soustraction d'arrière-plan se déroule généralement en deux phases. La première est la modélisation et la mise à jour de l'arrière-plan et la seconde est une différence pixel à pixel de l'image courante avec l'image arrière-plan construite [4,5,6]. Dans [7], les auteurs présentent les résultats d'évaluation des méthodes les plus connues que nous reprenons dans cet article afin de les tester sur nos images. La troisième approche est l'estimation de flot optique. Toutes les méthodes d'estimation du flot optique intègrent les informations sur un voisinage spatial et spatio-temporel pour estimer le mouvement apparent de l'objet [8].

[1] *http://www.irlynx.com/*

Dans ce papier, nous utilisons 5 méthodes de segmentation de la littérature auxquelles nous ajoutons notre proposition. Les algorithmes utilisés sont : l'algorithme statistique de modélisation d'arrière-plan (KDE) Kernel Density Estimation [7], l'algorithme de segmentation par seuillage, l'algorithme de soustraction d'arrière-plan basé sur l'aspect spatial et temporel [9] (STNBS), l'algorithme W^4*[10] qui est une version combinée de l'algorithme classique W^4 [5] avec la méthode de différence inter frame et enfin l'algorithme du filtre médian temporel qui est une approche de modélisation d'arrière-plan [6].

II. MÉTHODE PROPOSÉE

A. *Idée de l'approche*

Une analyse détaillée des images nous a permis de déterminer les facteurs qui permettront d'obtenir une bonne segmentation pour nos types de données.

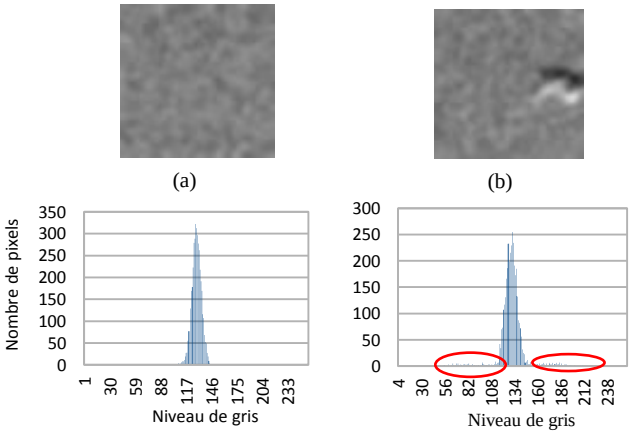


Figure 1. Répartition des niveaux de gris

En analysant les histogrammes des niveaux de gris des pixels (Fig.1-a) pour une image sans objet présent ou personne présente, on observe une concentration des niveaux de gris des pixels entre 100 à 150. Par contre il y a étalement des pixels sur les 255 niveaux de gris pour l'image contenant une personne (Fig.1-b). Cela nous montre que l'approche classique de segmentation par seuillage ou modélisation d'arrière-plan par une gaussienne peut permettre d'espérer séparer correctement l'objet et l'arrière-plan.

De plus, en considérant les informations spatiales des images, les attributs tels que l'écart type et l'étendue (écart entre l'intensité la plus élevée et l'intensité la moins élevée prises par les pixels de la région considérée) présentent des différences considérables d'une région de l'image à l'autre. Ces caractéristiques peuvent être exploitées pour accroitre la capacité de discrimination des régions de mouvement ou non et mieux classifier un pixel.

B. *Principe de l'algorithme*

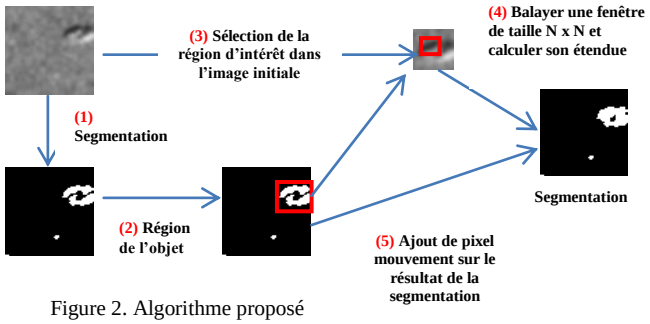


Figure 2. Algorithme proposé

L'approche que nous proposons (Fig.2) utilise les résultats d'algorithmes de segmentation pour améliorer la détection. L'amélioration consiste à reconstruire une forme, un contour plus ou moins identique à celui observé dans l'image initiale. Une correspondance de forme retrouvée avec celles d'une base d'apprentissage permettra de retrouver la posture de la personne. Pour parvenir à notre but, nous utilisons des informations spatiales d'une région d'intérêt pour ajouter des pixels mouvements. Notre but est d'identifier les pixels mouvements qui ne sont pas détectés avec les algorithmes classiques de segmentation. On constate en effet que les pixels se situant au centre de l'objet ont une valeur identique à celle de l'arrière-plan et il devient difficile algorithmiquement de caractériser ces pixels comme arrière-plan ou non (Fig.3). Par conséquent, en balayant une fenêtre entre la zone claire et sombre de l'image initiale, on a une forte probabilité d'obtenir un pixel clair et un autre sombre dans la sous-fenêtre et alors une valeur de l'étendue élevée. L'écart-type étant un critère pour caractériser la présence ou non de personne, sera aussi une variable pour atteindre notre objectif.

Les détails de l'approche sont donnés ci-dessous:

- première étape, nous appliquons un algorithme de segmentation qui, généralement, donne un résultat comme celui de la figure 3: un résultat dans lequel la forme et le contour ne sont pas retrouvés de manière nette.

Figure 3 : Exemple de résultat de segmentation obtenue à l'issue de l'étape 1

- deuxième étape, le résultat de segmentation précédent est utilisé pour récupérer les informations (la position, la taille) de la région où se trouve l'objet,
- troisième étape, on extrait de l'image initiale la sous fenêtre correspondant à cette région d'intérêt,
- quatrième et cinquième étape, à partir de la région sélectionnée, nous balayons une fenêtre de taille $N \times N$ dans l'image initiale. Pour chaque fenêtre nous calculons la valeur de son étendue. Cette valeur est comparée à l'écart-type des pixels de la fenêtre

considérée multiplié par un facteur k. Si $étendue > k\sigma$, le pixel central de la fenêtre est marqué comme étant un pixel mouvement sur le résultat de l'image segmentée (étape 5).

σ: représente l'écart-type des valeurs des pixels dans l'image

k: représente un paramètre de pondération.

III. EXPERIMENTATIONS ET RESULTATS

A. *Expérimentations*

Cinq algorithmes de segmentation de la littérature ont été implémentés et nous avons ajouté notre idée qui est celle de compléter les pixels mouvement en utilisant des informations spatiales. Par la suite ces algorithmes ont été comparés afin de trouver la meilleure approche de segmentation.

Les méthodes ont été implémentées en C++ avec la bibliothèque OpenCV sur un Core i7 Intel avec 16 Go de RAM.

1) *Choix des paramètres*

Dans cette étude, on choisit $N = 4$ comme taille de la fenêtre, la région d'intérêt fait environ 20x10 pixels. En prenant comme taille de la fenêtre glissante 4x4, celle-ci représente environ $1/10$ ième de la région d'intérêt. D'un point de vue détection d'objet, cette valeur de 4 semble la plus adéquate. En effet plus nous augmentons la taille, plus nous avons tendance à compléter l'ensemble des pixels de la région *(voir fig.4-c)*. Plus nous la diminuons *(voir fig.4-a)*, moins nous réussissons à retrouver une forme similaire à celle de la vérité terrain.

(a) N=3 (b) N=4 (c) N=5

Figure 4. Exemple de résultats de remplissage (en rouge) en variant la taille de la fenêtre pour k= 4

Pour le choix du paramètre k, nous avons évalué le pourcentage de classification correcte (PCC, cf III.B) pour différentes valeurs de k (Fig.5). Ce résultat montre que pour $k = 4$ nous obtenons le meilleur taux de classification correcte des pixels.

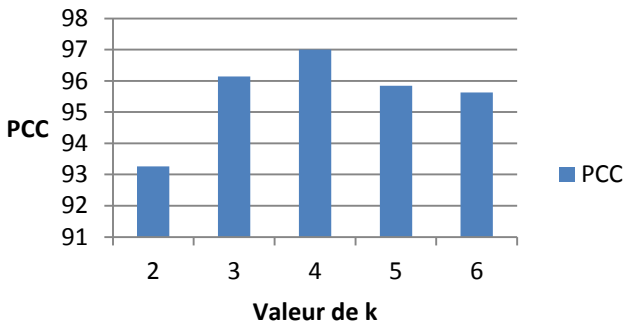


Figure 5: Pourcentage de classification correcte pour différentes valeurs de k

2) *Résultats*

Nous n'exposons dans cet article que les résultats obtenus pour la taille de la fenêtre = 4x4 et k=4. Nos expérimentations sont faites sur une vidéo contenant 200 frames. Dans cette vidéo, nous utilisons 20 frames : ceux dans lesquelles, on a une détection (variance >8). Nous avons choisi de présenter les résultats de 4 images consécutives dans lesquelles une personne se déplace. Ces images sont d'abord passées à un filtre gaussien pour un lissage, puis différents algorithmes de segmentation sont appliqués. Le tableau I regroupe les différents résultats obtenus.

B. *Evaluations*

Pour valider les résultats, des méthodes d'évaluation supervisée et non supervisée sont utilisées. Les résultats présentés dans le tableau 2 sont des moyennes de ceux obtenues sur 20 images qui ont été expertisées de façon à disposer d'une vérité terrain délimitant la zone en mouvement.

L'évaluation supervisée consiste à comparer la similarité du résultat de segmentation à celle de la vérité terrain. Notons toutefois que la vérité établie est en fonction des formes observées dans les images provenant du capteur et non celle d'une caméra du visible. Pour cette catégorie, nous appliquerons principalement 3 mesures à savoir F-mesure, le coefficient de Jaccard et le pourcentage de classification correcte.

Etant donné un résultat de segmentation et sa vérité terrain correspondante, on définit 3 grandeurs :

- la précision : $P = VP/(VP + FP)$
- le rappel : $R = VP/(VP + FN)$
- la spécificité : $S = VN/(VN + FP)$

(Avec VP : Vrai Positif, VN : Vrai Négatif, FP : Faux Positif, FN : Faux Négatif)

Les mesures d'évaluation utilisées sont alors :

- F-mesure : $F = 2(\frac{P*R}{P+R})$
- Coefficient de Jaccard : $JC = VP/(VP + FP + FN)$
- Pourcentage de classification correcte : $PCC = (VP + VN)$/(VP+FN+FP+VN)

L'évaluation non supervisée est intéressante dans notre contexte puisque nous travaillons sur des images particulières pour lesquelles il est difficile de dessiner la forme exacte de la personne. Ceci est due au fait que nous ne travaillons pas sur

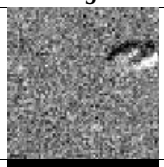
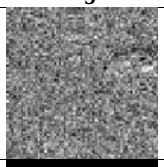
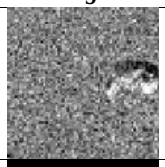
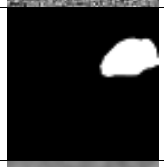
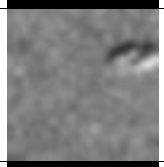
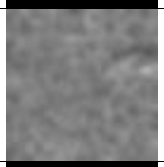
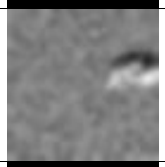
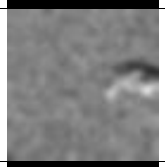
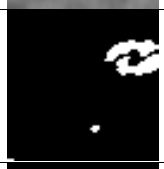
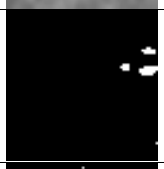
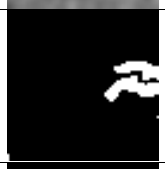
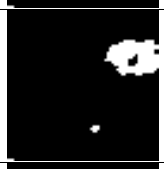
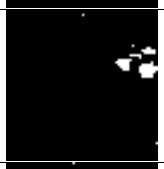
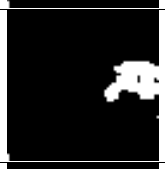
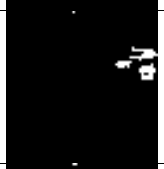
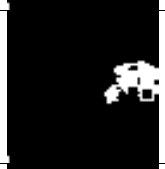
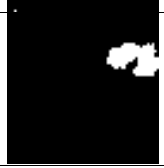
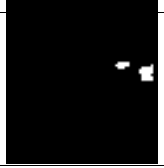
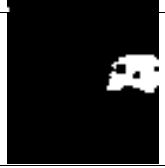
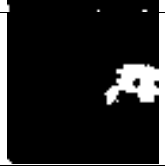
Algorithmes	***Images consécutives***			
	Image1	***Image2***	***Image3***	***Image4***
Image originale				
Vérité terrain				
Filtre gaussien				
Exemple de segmentation obtenue à l'issue de l'étape 1				
Seuillage				
KDE				
STNBS				
W4*				
Median				

Tableau I : Images traitées et segmentations finales obtenues par les différents algorithmes modifiés

des images du visible. Pour l'évaluation non supervisée, on considère l'image initiale et le résultat de la segmentation et on applique la méthode intra-inter région proposée dans [11] et implémentée dans [12]. Cette méthode consiste à calculer la somme des contrastes des régions pondérées par leurs aires et la somme des variances normalisées des régions. Nous avons d'abord comparé les résultats de segmentation des algorithmes tels que proposés par les auteurs (non supervisé sans remplissage) et ensuite nous avons comparé ceux dans lesquels on complète les pixels qui constituent l'objet segmenté (non supervisé avec remplissage).

Algorithme	**Supervisé**			**Non supervisé (sans remplissage)**	**Non supervisé (avec remplissage)**
	F	***JC***	***PCC***	***Intra_Inter***	***Intra_Inter***
Seuillage	0,747	0,627	0,970	0,623	0,595
KDE	0,757	0,634	0,970	0,608	0,594
STNBS	0,721	0,591	0,961	0,605	0,572
W4*	0,668	0,517	0,957	0,523	0,503
Médian	0,662	0,535	0,964	0,579	0,568

Tableau II : Résultats d'évaluation

C. Interprétations

D'après les résultats de l'évaluation supervisée, notre approche combinée avec KDE et l'approche combinée avec le seuillage présentent de meilleurs résultats comparés aux autres méthodes. Nous pouvons voir (Tableau II) que nous avons un pourcentage de classification correcte des pixels de 97%. On note aussi une légère différence de 1% entre KDE et seuillage pour des critères d'évaluation F-mesure et distance de Jaccard. Par contre pour la segmentation non supervisée les résultats avec seuillage sont meilleurs que toutes les autres méthodes.

Notons aussi que l'approche de post-traitement proposée met environ 2 ms d'exécution par image. Pour 20 images, cette approche combinée avec l'algorithme de segmentation par seuillage, met 57ms de temps total d'exécution. Cela montre que nous restons bel et bien dans un cas de détection en temps réel. Le tableau III présente les temps d'exécution des algorithmes de segmentation combinés avec la méthode proposée.

Algorithme	**Seuillage**	**KDE**	**STNBS**	**W4***	**Médian**
Temps total d'exécution (millisecondes)	**57**	290	139	146	105

Tableau III : Temps d'exécution des algorithmes

IV. CONCLUSION ET PERSPECTIVES

Nous avons présenté une approche de segmentation adaptée au capteur infrarouge. Cette approche a pour but de retrouver la forme et le contour tels qu'observés dans l'image initiale. Elle se base sur un résultat de segmentation et aussi exploite des informations spatiales principalement les attributs du premier ordre pour mieux segmenter l'objet. D'après l'étude menée, il en ressort que l'approche combinée avec un résultat de segmentation par seuillage ou par KDE présente les meilleurs résultats. L'approche proposée ici sera utilisée dans les travaux futurs pour la reconnaissance d'activités ou de postures afin d'avoir une méthode complète pouvant être implémentée dans les solutions d'assistance aux personnes âgées. Cette approche pourra être adaptée pour la détection, la surveillance et la sécurité dans des scènes (chambre, bureau, bâtiment du tertiaire ...). Le type de données sur lequel nous travaillons étant particulier, pour la suite de nos travaux nous proposons de constituer une base de données publique et nous continuerons le processus d'analyse comportementale par les aspects d'analyse de posture (assise, couchée, debout …) et d'intensité d'activité.

REMERCIEMENTS

Les auteurs tiennent à remercier BPI France, les conseils régionaux du Limousin et de Rhône-Alpes associé au FEDER, le conseil départemental de l'Isère, et la communauté d'agglomération Bourges Plus, pour leur soutien financier au projet CoCAPs. Le projet CoCAPs, issu du FUI N°20, est également soutenu par les pôles de compétitivité S2E2, Minalogic.

REFERENCES

[1] S. Vishwakarma et A. Agrawal, "A survey on activity recognition and behavior understanding in video surveillance", *The Visual Computer*, vol. 29, n° 10, p. 983-1009, 2013.

[2] Collins, Robert T., Lipton, Alan J., KANADE, Takeo, *et al.* "A system for video surveillance and monitoring". 2000.

[3]Kameda, Yoshinari et Minoh, Michihiko. "A human motion estimation method using 3-successive video frames". In : *International conference on virtual systems and multimedia*. 1996. p. 135-140.

[4] Elgammal Ahmed, Duraiswami Ramani, Harwood David, ." Background and foreground modeling using nonparametric kernel density estimation for visual surveillance". *Proceedings of the IEEE*, 2002, vol. 90, no 7, p. 1151-1163.

[5] Haritaoglu Ismail, Harwood David, et Davis Larry S.. . "W^4: real-time surveillance of people and their activities". *IEEE Transactions on pattern analysis and machine intelligence*, 2000, vol. 22, no 8, p. 809-830.

[6] Hung Mao-Hsiung, Pan Jeng-Shyang et Hsieh Chaur-Heh. "A fast algorithm of temporal median filter for background subtraction". *Journal of Information Hiding and Multimedia Signal Processing*, 2014, vol. 5, no 1, p. 33-40.

[7] M. Hedayati, W. M. D. W. Zaki, et A. Hussain, "Real-time background subtraction for video surveillance: From research to reality", in *2010 6th International Colloquium on Signal Processing and Its Applications (CSPA)*, 2010, p. 1-6.

[8] Barron John L., Fleet David J., et Beauchemin, Steven S." Performance of optical flow techniques". *International journal of computer vision*, 1994, vol. 12, no 1, p. 43-77.

[9] Shengping Zhang, Hongxun Yao, et Shaohui Liu. 2009. " Spatial-temporal nonparametric background subtraction in dynamic scenes". In , 518-21. IEEE. doi:10.1109/ICME.2009.5202547.

[10] Yin Jiale, Liu Lei, Li He. "The infrared moving object detection and security detection related algorithms based on W4 and frame difference". *Infrared Physics & Technology*, 2016, vol. 77, p. 302-315.

[11] Levine Martin D. et Nazif Ahmed M. "Dynamic measurement of computer generated image segmentations". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1985, no 2, p. 155-164.

[12] S. Chabrier, B. Emile, C. Rosenberger, H. Laurent, "Unsupervised performance evaluation of image segmentation", *Special Issue on Performance Evaluation in Image Processing, EURASIP Journal on Applied Signal Processing*, pages 1-12, 2006