Implémentation optimisée d'un classifieur neuronal pour la détection en temps réel de personnes à terre

Olivier Boisard[1,2], Guillaume Sauvage[1], Olivier Brousse[1,2], Michel Paindavoine[2]

[1]GlobalSensing Technologies
14 rue Pierre de Coubertin, bât. I, 21000 Dijon

[2]Université de Bourgogne, LEAD, CNRS UMR 5022 Pôle AAFE, 11 Esplanade Erasme, 21000 Dijon
olivier_boisard@etu.u-bourgogne.fr, guillaume.sauvage@globalsensing.eu
olivier.brousse@globalsensing.eu, paindav@u-bourgogne.fr

Résumé – Cet article présente un système de détection quasi-temps réel de personnes tombées à terre, basé sur un capteur d'images CMOS standard. Il consiste à supprimer le fond de l'image pour extraire les objets en mouvement, et à créer des boîtes englobantes autour de ces objets ; selon leurs orientations, le système décide s'il y a une personne à terre ou non. Afin de fiabiliser le système et d'éviter les fausses alarmes, nous proposons d'utiliser un algorithme neuro-inspiré pour nous assurer que l'objet détecté est bien une personne.

Abstract – This paper presents an almost-real-time fallen person detection system, based on standard CMOS image sensor. The process first suppresses the background in order to extract moving objects, and creates bounding boxes around them. The orientations of those bounding boxes allow us to know whether there is a fallen person. To strengthen the system and avoid false alarms, we use a neuro-inspired algorithm that checks that the detected objects are human beings.

1 Introduction

Une des principales applications de la domotique est la sécurité de l'habitant dans son domicile. En ce qui concerne les personnes âgées, la chute représente un des principaux dangers. En effet, en France, un tiers des personnes âgées de plus de 65 ans chutent en moyenne une fois dans l'année. Après 80 ans, c'est une personne sur deux qui est concernée. Pour cette population, la chute représente la principale cause de décès. On estime à 2,7 millions le nombre de chutes en France pour la seule année 2004 [1, 5, 6]. Pour limiter les risques de décès ou de séquelles, il est essentiel d'intervenir rapidement ; il est donc nécessaire d'imaginer des systèmes capables de détecter ces chutes, et de prévenir l'entourage de la personne. Différentes approches sont possibles [4]. Tout d'abord, la personne peut porter sur elle un dispositif muni d'un bouton, qu'elle presse en cas de problème. Ces systèmes ont l'avantage d'être extrêmement simples et bon marchés, mais présentent deux inconvénients majeurs : la personne doit être consciente, ce qui n'est pas toujours le cas après une chute ou si la chute est due à un malaise, et elle doit l'avoir sur elle en permanence – certaines personnes oublient, ou même refusent de porter ce genre de système qu'elles peuvent trouver stigmatisants. Une autre approche consiste à demander à la personne de porter un système intelligent, capable de détecter la chute. Si cette approche permet de s'affranchir de l'état de conscience de la personne, il reste le problème de refus ou d'oubli. Enfin, une troisième possibilité consiste à équiper l'habitat de capteurs. Cette approche est probablement la plus fiable ; en effet, elle ne dépend pas de la personne. Son seul inconvénient est qu'elle peut donner l'impression de « surveiller »la personne dans son propre domicile. Cependant, les capteurs peuvent être suffisamment bien cachés dans l'habitat pour atténuer ce phénomène. L'approche que nous proposons dans cet article appartient à cette troisième catégorie de systèmes. Elle est basé sur un capteur CMOS standard et sur une chaîne de traitement, fiabilisée grâce à un algorithme neuro-inspiré.

Certains détecteurs de chutes basés sur le traitement vidéo analysent les changement dans la posture de l'individu [2, 7]. Dans le cas d'une chute, ces changements sont très rapides, ce qui nécessite un traitement temps réel complexe ; bien que performants, ces systèmes demandent une grande puissance de calcul, ce qui augmente leur coût de revient. L'approche inverse consiste à proposer un système moins performant, mais également moins consommateur et moins cher [12, 11].

Nous proposons une approche « hybride » : notre système ne requiert pas de machine particulièrement coûteuse, mais nous exécutons malgré tout des algorithmes complexes pour fiabiliser les détections. Pour cela, nous ne détectons pas la chute elle-même, mais sa conséquence :

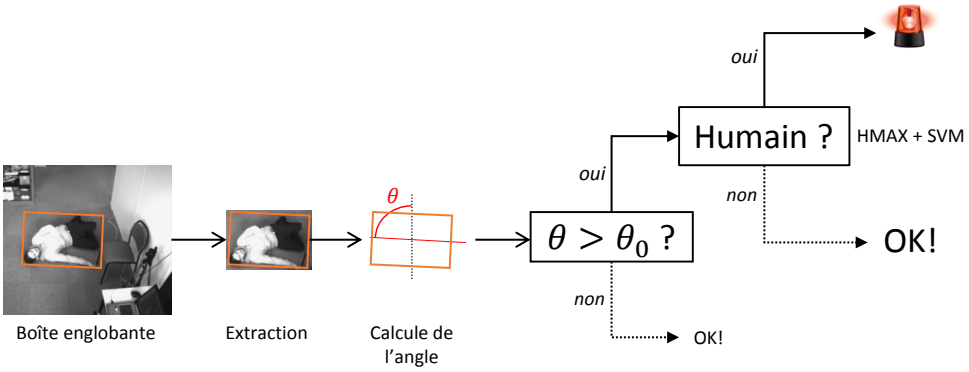

FIG. 1: Chaîne de traitements pour la détection de personnes à terre

nous levons une alarme lorsque nous détectons une personne à terre. Ceci nous autorise à traiter les images à une fréquence bien inférieure à 25 images par secondes.

Dans cet article, nous allons tout d'abord présenter en détail notre système en Section 2. La Section 3 sera consacrée à la validation de notre système, et à sa comparaison avec les systèmes actuels. Enfin, les conclusions de ces travaux et les améliorations envisagées seront présentées en Section 4.

2 Notre système

2.1 Présentation

Le principe de fonctionnement de notre système est semblable à celui proposé par Williams *et al* [12]. Il consiste à extraire les personnes de l'image avec un algorithme de suppression de fond basé sur la détection de mouvement, puis à définir des boîtes englobantes autour d'elles. Elle prennent la forme de rectangles, définis par une largeur L, une hauteur H vérifiant $H > L$, et un angle θ non orienté avec l'axe vertical. Si cet angle est inférieur à un seuil θ_0, nous considérons que la personne est debout. Si nous avons $\theta > \theta_0$, nous considérons que la personne est à terre, et qu'une alarme doit être levée.

En revanche, l'algorithme de suppression de fond fait ressortir tous les objets en mouvements, et pas seulement les personnes. Cela peut amener l'algorithme de boîtes englobantes à détecter d'autres objets, par exemple une chaise que l'on déplace, ou un manteau qui tombe, ce qui peut avoir pour conséquence un fort taux de faux positifs et rendre le système pénible pour l'utilisateur. Afin de limiter ce problème, nous proposons de vérifier la nature de l'objet détecté par l'algorithme de boîtes englobantes. Pour cela, nous utilisons la chaîne de traitement HMAX, créée par Serre *et al* [9] et qui est un modèle du système de vision chez les primates. Ce modèle est constitué de quatre couches S1,C1,S2 et C2. S1, est composée d'une batterie de filtres de Gabor appliqués à l'image pour en extraire des caractéristiques ; C1, qui sous-échantillonne les images en sortie de S1 avec des maximas locaux ; S2, qui compare les sorties de C1 avec un ensemble d'images pré-apprises ; et C2, qui ne conserve que les sorties maximums de S2. Ces sorties sont ensuite envoyées à un classificateur SVM, qui déterminera si l'image d'entrée était celle d'un humain ou non. Les performances d'HMAX en termes de classification d'images sont très élevées, mais son inconvénient majeur et qu'il est, d'un point de vue algorithmique, hautement complexe. En réponse à ce problème, nous avons implanté les optimisations proposées par Yu et Slotine [13], qui consiste entre autre à remplacer la batterie de filtres de Gabor de l'étage S1 par une décomposition en ondelettes. Nous avons choisi d'utiliser des ondelettes de Haar, mais il est également possible d'implanter de manière optimisée d'autres types d'ondelettes sur des systèmes embarqués [3]. La chaîne de traitement complète est illustrée en Figure 1. Nous utilisons également la méthode de Yu et Slotine pour réduire le nombre de vecteurs du dictionnaire de S2 [13], et ainsi grandement accélérer le traitement.

2.2 Analyse de complexité

Analysons maintenant la complexité algorithmique de notre système. La première étape est donc l'extraction de boîtes englobante ; cependant, la complexité de cet algorithme est négligeable en regard de celle du modèle HMAX. Il en va de même pour le classificateur SVM à la fin du traitement.

La couche S1 de HMAX consiste en une décomposition en ondelettes de Haar à 3 échelles. Pour une image de dimmensions $W \times H$. Le nombre d'opérations s'élève à :

$$C_{\mathrm{S1}} = \frac{9}{4} WH$$

La couche C1, quand à elle, consiste un en filtre par maxima locaux, avec des fenêtres de 2×2 sans recouvrement. Appliqué aux images en sortie de S1, cela représente un total d'opérations C_{C1}, avec :

$$C_{\mathrm{C1}} = \frac{63}{256} WH$$

La couche S2 est la plus complexe – cependant, il est possible de l'optimiser en passant par des convolutions, comme cela est réalisé dans l'implémentation proposé par Serre *et al* [9]. Pour chaque patch de $W_p \times H_p \times 3$ pixels, cela représente un total de $C\left(S_{2p}\right) = \frac{63}{256} W H W_p H_p$ opérations. En considérant N_{S2} patches de dimensions moyennes $\bar{W}_p \times \bar{H}_p$, cela représente un total d'opérations qui s'élève à :

$$C_{\mathrm{S2}} = \frac{63}{256} N_{\mathrm{S2}} W H \bar{W}_p \bar{H}_p$$

Enfin la couche C2, qui consiste à ne prendre que les sorties maximums pour chaque patch de S2, prend C_{C2} opérations, avec :

$$C_{\mathrm{C2}} = \frac{21}{256} W H N_{\mathrm{S2}}$$

Supposons qu'après avoir découpé la région de l'image délimitée par la boîte englobante, cette région soit redimensionnée en 140×280 pixels avant d'être traité par HMAX, et que le dictionnaire de donnée de la couche S2 comporte 200 patches de dimension moyennes 10×10. Avec ces paramètres, les contributions de chacune des couches à la complexité de l'ensemble sont données en Tableau 1. On voit qu'en comparaison de la couche S2, les complexités algorithmiques des autres couches sont négligeables – à peine 0,4 % à elles trois. Par conséquent, nous pouvons considérer :

$$C_{\mathrm{HMAX}} \approx C_{\mathrm{S2}} \approx 193 \text{ Mops.}$$

Ces opérations peuvent être facilement parallélisées : en effet, chaque traitement impliquant un patch du dictionnaire de S2 est indépendant des autres. Cela signifie que ce traitement pourrait être lancer dans 200 processus parallèles, par exemple sur un GPU ou sur une architecture dédiée.

TAB. 1: Complexité des différentes couches de HMAX

C_{S1}	0,04553939 %
C_{C1}	0,00498087 %
C_{S2}	99,61742166 %
C_{C2}	0,33205807 %

3 Expérimentations et comparaisons aux autres systèmes

Nous avons enregistré 20 séquences vidéos d'une personne évoluant dans un intérieur dégagé, comme montré en Figure 1. Sur 10 de ses séquences, la personne réalise des tâches de la vie de tous les jours : marcher, s'asseoir, se lever, etc. Sur les 10 autres séquences, la personne réalise les même tâche, mais cette fois-ci en incluant une chute.

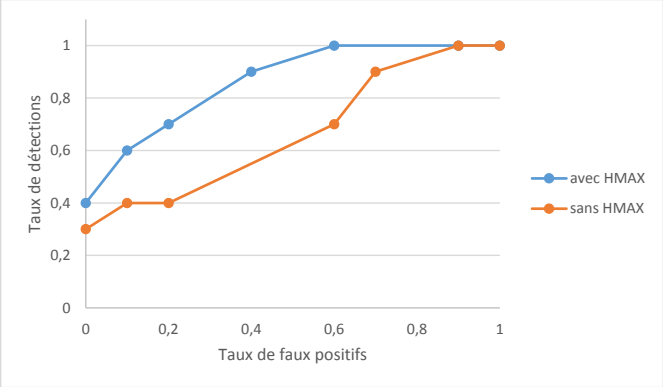

FIG. 2 : Taux de détections en fonction du taux de faux positifs.

Nous avons appliqué un filtrage temporel en guise de post-traitement : une alarme ne sera levée que si une personne à terre est détectée sur un certains nombre N d'images consécutives, éliminant ainsi la plupart des faux positifs. Plus N sera grand, moins le système sera sensible. La Figure 2 présente le taux de détections en fonction du taux de faux positifs.

Nous avons comparé notre système avec la pointe de ce qui existe actuellement [4]. Nous ne comparons notre système qu'à ceux qui lui sont comparables, c'est-à-dire ceux basés sur une caméra vidéo intégrée dans l'habitat. Le Tableau 2 résume leurs caractéristiques [4]. Le système de Chen *et al* [2] identifie les variations de postures, et exige donc une machine suffisemment puissante pour assurer un traitement à une fréquence suffisante. Notre système ne présente pas cet inconvénient, puisqu'il détecte non-pas un mouvement, mais une posture. Humenberger *et al* ont proposé d'utiliser une caméra 3D, c'est-à-dire du matériel spécifique ; dans notre cas, une simple caméra basée sur un capteur CMOS suffit. Shoaib *et al* utilisent une analyse de contexte complexe, là ou notre système se contente d'opérer une suppression de fond – tous les autres traitement se font image par image, de manière indépendante. Enfin, aucun de ces systèmes ne vérifie que l'objet détecté est bien humain, ce qui peut générer un grand nombre de faux positifs, en particulier si la personne à des animaux. En considérant que les performances générales peuvent être évaluée en calculant l'aire sous les courbes présentées en Figure 2, le fait d'utiliser HMAX nous permet de passer de 64 % de performances à 80 %, soit un gain de 16 point. L'étude de complexité montre que cela ajoute certes un nombre non-négligeable de calculs ; néanmoins cela ne pose pas de problèmes dans le cas où l'on cherche à détecter une personne à terre, et non en train de tomber. Avec un processeur cadencé à 1 GHz, il est possible de détecter de l'ordre d'une personne à terre par image, ce qui reste acceptable dans ce cadre.

TAB. 2: Comparaison de notre système avec l'existant

Référence	Consommation	Coût	Complexité algorithmique	Précision
Chen [2]	élevée	élevé	élevée	90,09 %
Yu [14]	élevée	élevé	élevée	97,08 %
Humenberger [8]	élevée	élevé	élevée	90 % - 99 %
Shoaib [10]	élevée	moyen	élevée	96 %
Notre système	**élevée**	**moyen**	**élevée**	**80 %**

4 Conclusion

Dans cet article, nous avons proposé un moyen de fiabiliser un système de chute de personne, basé sur une caméra vidéo. Le système consiste, après détection d'une chute, à vérifier que l'objet ayant chuté est bien une personne grâce à un algorithme neuro-inspiré, afin d'éliminer les fausses alertes. Cela se traduit certes par une complexité algorithmique plus élevée, mais cela nous a permis de grandement réduire le taux de faux positifs. Les travaux futures viseront à optimiser l'algorithme de classification neuro-inspiré, afin de réduire la complexité algorithmique de l'ensemble, et donc ainsi son prix de reviens et sa consommation en énergie.

Références

[1] Haute autorité de santé. Évaluation et prise en charge des personnes âgées faisant des chutes répétées. 2009.

[2] Y.-T. Chen, Y.-C. Lin, and W.-H. Fang. A hybrid human fall detection scheme. In *2010 17th IEEE International Conference on Image Processing (ICIP)*, pages 3485–3488, September 2010.

[3] S. Courroux, S. Chevobbe, M. Darouich, and M. Paindavoine. Use of wavelet for image processing in smart cameras with low hardware resources. *Journal of Systems Architecture*, 59(10) :826–832, November 2013.

[4] Y.-S. Delahoz and M.-A. Labrador. Survey on Fall Detection and Fall Prevention Using Wearable and External. *Sensors*, 14(10) :19806–19842, October 2014.

[5] DREES. L'état de santé de la population en france. 2011.

[6] Réseau francophone de prévention des traumatismes et de promotion de la sécurité. Prévention des chutes chez les personnes âgées à domicile. 2005.

[7] Y. Hirata, A. Muraki, and K. Kosuge. Motion control of intelligent passive-type Walker for fall-prevention function based on estimation of user state. In *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006*, pages 3498–3503, 2006.

[8] M. Humenberger, S. Schraml, C. Sulzbachner, A.-N. Belbachir, Á. Srp, and F. Vajda. Embedded fall detection with a neural network and bio-inspired stereo vision. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Providence, RI, USA, June 16-21, 2012*.

[9] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio. Robust object recognition with cortex-like mechanisms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(3) :411–426, 2007.

[10] M. Shoaib, R. Dragon, and J. Ostermann. View-invariant fall detection for elderly in real home environment. In *2010 Fourth Pacific-Rim Symposium on Image and Video Technology (PSIVT)*, pages 52–57, November 2010.

[11] V. Vaidehi, K. Ganapathy, K. Mohan, A. Aldrin, and K. Nirmal. Video based automatic fall detection in indoor environment. In *2011 International Conference on Recent Trends in Information Technology (ICRTIT)*, pages 1016–1020, June 2011.

[12] A. Williams, D. Ganesan, and A. Hanson. Aging in Place : Fall Detection and Localization in a Distributed Smart Camera Network. In *Proceedings of the 15th International Conference on Multimedia*, MULTIMEDIA '07, pages 892–901, New York, NY, USA, 2007. ACM.

[13] G. Yu and J.-J. Slotine. FastWavelet-Based Visual Classification. In *19th International Conference on Pattern Recognition, 2008. ICPR 2008*, pages 1–5, 2008.

[14] M. Yu, A. Rhuma, S.M. Naqvi, L. Wang, and J. Chambers. A posture recognition-based fall detection system for monitoring an elderly person in a smart home environment. *Information Technology in Biomedicine, IEEE Transactions on*, 16(6) :1274–1286, November 2012.