

Analyse Spatiotemporelle des Vecteurs de Mouvement : Application au Comptage des Personnes

Yassine Benabbas*, Tarek Yahiaoui*, Thierry Urruty*, Chabane Djeraba*

*LIFL UMR CNRS 8022, Université de Lille1, TELECOM Lille1
IRCICA, Parc scientifique de la Haute Borne, 56950 Villeneuve d'Ascq
{yassine.benabbas,tarek.yahiaoui,thierry.urruty,chabane.djeraba}@lifl.fr

Résumé. Cet article présente une nouvelle approche qui permet de compter le nombre d'individus franchissant une ligne de comptage. L'approche proposée accumule dans le temps les vecteurs de mouvement pour chaque point de la ligne de comptage formant une carte spatiotemporelle. Une procédure de détection en ligne des blobs est ensuite utilisée afin de déterminer les régions de la carte spatiotemporelle qui correspondent à des personnes franchissant cette ligne. Le nombre d'individus associé à chaque blob est estimé grâce à un modèle de régression linéaire appliqué aux caractéristiques du blob. L'approche proposée est validée sur la base de plusieurs ensembles de données enregistrées à l'aide d'une caméra verticale ou d'une caméra oblique.

1 Introduction

Le comptage des personnes est une application importante dans divers domaines allant du marketing à la vidéosurveillance. Lorsque de nombreuses personnes traversent le hall d'une gare, d'un bâtiment ou d'un centre commercial, leur nombre représente une information pertinente pour les décideurs chargés d'assurer la sécurité publique et d'élaborer des stratégies commerciales. Ainsi, un certain nombre de personnes dans un contexte particulier est susceptible de refléter une situation anormale et potentiellement dangereuse. De plus, la possibilité de connaître le nombre d'individus dans une zone commerciale permet d'obtenir de précieuses informations pour les gérants d'un magasin afin d'évaluer l'attractivité des offres promotionnelles qu'ils proposent.

De nombreuses approches de comptage ont été proposées dans la littérature (Djeraba et al., 2010). Le problème est souvent simplifié par l'utilisation d'une caméra zénithale (verticale) (Antic et al., 2009), une caméra frontale (Zhao et al., 2009) ou une configuration multi-caméras (Yang et al., 2003). On peut classer les méthodes utilisant une seule caméra en deux catégories :

Méthodes basées sur la détection des personnes : plusieurs algorithmes de détection de personnes ont été proposés ces dix dernières années (Enzweiler et Gavrilu, 2009). Généralement, lorsque les individus sont face à la caméra, des algorithmes de détection frontale du visage (Chen et al., 2009) sont utilisés. Cependant, ces méthodes montrent leurs limites dans des situations de foule ou lorsque des occlusions surviennent.

Méthodes basées sur les caractéristiques : au lieu de détecter des individus, ces méthodes extraient diverses caractéristiques pour effectuer le comptage. Elles sont aujourd'hui devenues

Comptage des Personnes

populaires car elles permettent de préserver l'anonymat des personnes (Chan et al., 2008). On trouve dans cette catégorie les méthodes spatio-temporelles (Cong et al., 2009) qui construisent une carte spatio-temporelle en empilant chaque ligne de comptage à travers le temps. Des modèles statistiques sont ensuite utilisés pour définir le nombre de personnes franchissant la ligne. Ces méthodes sont simples et rapides à appliquer mais ne gèrent pas certains cas qui surviennent en pratique (par exemple, lorsqu'une personne s'arrête sur la ligne de comptage, un blob de taille très importante est généré et incorrectement assimilé à un groupe de personnes).

Notre approche permet de compter le nombre d'individus franchissant une ligne de comptage à partir de vidéos issues de caméras monoculaires. Elle fait partie des approches spatio-temporelles avec l'originalité d'éviter la détection des individus qui s'arrêtent sur la ligne de comptage en utilisant les vecteurs de flux optique. Ces vecteurs permettent également d'obtenir l'orientation du blob de manière plus efficace. Notre approche est capable de détecter en temps réel les blobs qui franchissent la ligne contrairement aux approches précédentes qui attendent la construction complète de la carte spatio-temporelle avant d'entamer la détection des blobs. L'utilisation d'un modèle de régression linéaire pour chaque angle de prise de vue permet à notre approche d'être indépendante de l'angle de prise de vue.

2 Notre approche

Nous proposons une approche permettant de compter les personnes franchissant une ligne de comptage et pouvant être adaptée à différentes configurations d'installation. La Figure 1 montre deux exemples de configuration de caméra pour lesquels notre approche est adaptée.



FIG. 1 – Exemples de configuration de la caméra : (a) Caméra orientée verticalement au-dessus de la tête des passants, (b) Caméra orientée obliquement.

Notre approche se divise en deux phases comme indiqué en Figure 2. La première *extraction des blobs* comporte quatre étapes pour extraire les blobs correspondant aux personnes franchissant la ligne de comptage. La phase *décision de comptage* comporte deux étapes associant le nombre de personnes dans chaque blob et leur direction.

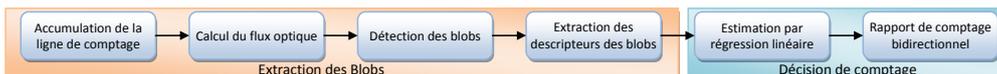


FIG. 2 – Architecture du système proposé.

2.1 Extraction des blobs

Nous définissons d'abord une ligne d'intérêt dans la vidéo appelée "ligne de comptage" et nous analysons les changements qui y interviennent à travers le temps. Les pixels se trouvant sur cette ligne forment une colonne dans la carte spatio-temporelle. Les personnes passant sur la ligne de comptage forment des zones spatialement et temporellement connexes qui s'empilent image par image comme illustré dans la Figure 3. La carte spatio-temporelle colorimétrique ainsi obtenue représente les variations des couleurs des pixels sur la ligne de comptage à travers le temps. Cette étape est encore illustrée par la Figure 4 où la carte spatio-temporelle est obtenue par l'accumulation de la ligne de comptage (représentée en rouge) sur une vidéo de la base PETS2009¹.

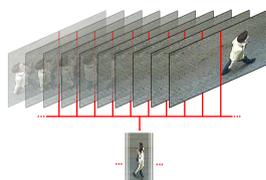


FIG. 3 – Accumulation de la ligne de comptage à travers le temps.



FIG. 4 – Carte spatio-temporelle obtenue par accumulation de la ligne de comptage en rouge dans une séquence vidéo.

L'étape suivante consiste à calculer les vecteurs de flux optique des pixels appartenant à la ligne de comptage pour chaque image de la séquence vidéo. Pour cela, nous utilisons la méthode de Lucas et Kanade (Lucas et Kanade, 1981). L'orientation et la vitesse des vecteurs obtenus sont empilées à travers le temps afin de construire des cartes spatio-temporelles d'orientation et de vitesse. Ces deux cartes sont utilisées lors de l'étape de détection en ligne des blobs pour détecter et mettre à jour leurs caractéristiques (position, vitesse, orientation et taille) jusqu'à ce que les blobs prennent une forme définitive (i.e. que la ou les personnes aient complètement franchi la ligne de comptage). Dans notre approche, nous utilisons les notations suivantes : L est la ligne de comptage de longueur l . $p_{i,t}$ est un pixel de la carte spatio-temporelle Map_L construite à partir de la ligne L à l'instant t avec $i = \overline{0..l-1}$. $OF(p_{i,t})$ est le vecteur de flux optique ayant pour origine le pixel $p_{i,t}$. L'ensemble des blobs détectés est noté S . Un blob identifié par Id est défini par le vecteur suivant :

$$B(Id) = (P, N, \alpha, \beta, l, h, O, V) \quad (1)$$

où P est l'ensemble des pixels du blob et N leur nombre. l et h sont respectivement la largeur et la hauteur du rectangle minimum englobant les pixels du blob. α et β sont les coordonnées du coin supérieur gauche de ce rectangle. O correspond à l'orientation moyenne du blob et V à sa vitesse moyenne.

L'algorithme de détection en ligne des blobs (Benabbas et al., 2010) est utilisé après le calcul des vecteurs de flux optique des pixels de la ligne de comptage pour l'image courante. Il prend uniquement en compte les pixels dont le flux optique a une vitesse non nulle pour filtrer les personnes statiques. Les pixels voisins de la colonne courante dans la carte spatio-temporelle ayant une orientation similaire sont regroupés dans le même blob. Les blobs ob-

1. <http://www.cvg.rdg.ac.uk/PETS2009/index.html>

tenus à partir de la colonne courante sont regroupés avec ceux obtenus à partir des itérations précédentes de l'algorithme. Lorsqu'un nouveau pixel $p_{i,t}$ est ajouté au blob $B(Id)$, alors l'algorithme met les paramètres du blob à jour.

L'étape d'extraction des descripteurs des blobs prend uniquement en compte ceux qui ont franchi la ligne de comptage et leur assigne les caractéristiques décrites dans la Formule 1. Ces descripteurs sont utilisés comme entrées pour la phase *décision de comptage*.

2.2 Décision de comptage

Nous utilisons un modèle de régression linéaire *pace regression* pour estimer le nombre d'individus dans un blob qui est donné par la formule suivante :

$$y = H(x) = \sum_{i=0}^n \theta_i x_i = \theta^T x \quad (2)$$

où y correspond à l'estimation du nombre d'individus dans un blob, et l'hypothèse $H(x)$ correspond au modèle de régression linéaire. n représente le nombre de caractéristiques d'un blob tel qu'indiqué dans la Formule 1. x est un vecteur de n dimensions contenant les caractéristiques du blob. Le vecteur de n dimensions des coefficients θ est estimé par un apprentissage supervisé. Nous avons adopté une méthode d'apprentissage hors ligne pour chaque configuration différente en utilisant une vérité-terrain annotée manuellement par un expert. L'utilisation d'un vecteur de coefficients distinct pour chaque configuration permet de supporter facilement de nouvelles configurations. Le nombre de personnes comptées dans le blob ainsi que son orientation sont ensuite rapportés à l'utilisateur en temps réel. La Figure 5 illustre les étapes clés de notre approche.

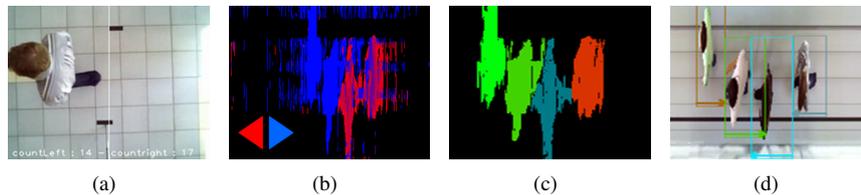


FIG. 5 – *Étapes-clés de notre approche : (a) Sélection de la ligne de comptage, (b) Carte spatio-temporelle des orientations, (c) Détection en ligne des blobs, (d) Représentation de l'orientation des blobs sur la carte spatio-temporelle.*

3 Expériences et résultats

Notre approche a été testée sur deux bases de vidéos. La première a une durée totale de 30 minutes tandis que la seconde a une durée totale de 2 heures. Le nombre de personnes pour les deux ensembles réunis dépasse les 1 000 personnes. Le Tableau 1 indique la précision de notre approche pour les séquences des deux bases de vidéos. Nous avons aussi testé l'approche sur une vidéo issue de la base PETS2009 (Figure 6). Cette figure représente une carte spatio-temporelle obtenue à partir de la ligne de comptage représentée en rouge dans la Figure 4. Chaque blob détecté est encadré par un rectangle contenant une flèche indiquant sa direction et un nombre représentant le comptage estimé. Le graphique montre la vérité-terrain et les résultats de notre approche pour chaque image.

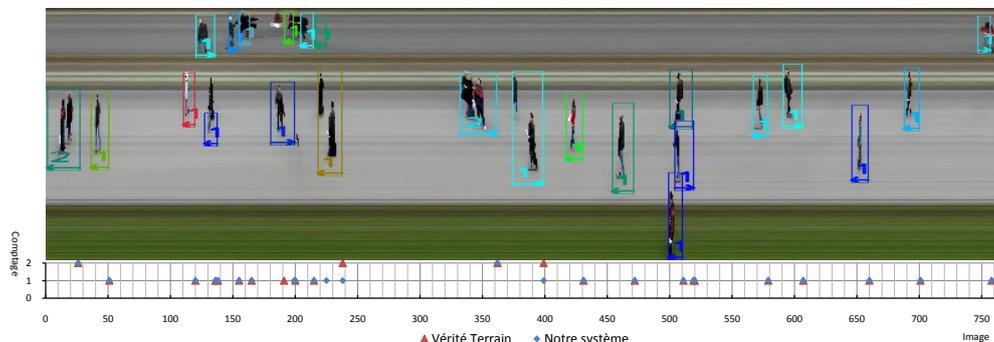


FIG. 6 – Résultats du comptage dans une séquence vidéo issue de la base PETS2009.

Id	Direction	Vérité terrain	Notre système	Précision
1-6	(0) Gauche	426	402	94.37%
	(1) Droite	233	236	98.71%
	(2) Les deux	659	638	96.81%
7-11	(0) Entrée	225	231	97.33%
	(1) Sortie	108	124	85.19%
	(2) Les deux	333	335	93.39%

TAB. 1 – Précision globale du système de comptage.

Approche	Processeur	FPS
Notre approche	Celeron 1.86GHz	45
(Antic et al., 2009)	3.0GHz	25
(Cong et al., 2009)	Pentium 2.8GHz	25
(Velipasalar et al., 2006)	Pentium 2 GHz	33

TAB. 2 – Précision globale du système de comptage.

Les résultats de la table 1 montrent que notre système fait preuve d'une grande robustesse avec une précision globale de 96.81% pour le premier ensemble de vidéos, et de 93.39% pour le second. La grande différence entre $A_{7-11,0} = 97.33\%$ et $A_{7-11,1} = 85.19\%$ s'explique par le fait que les clients présents dans nos vidéos ont généralement rempli leurs sacs avant de quitter le magasin, ce qui génère des blobs plus grands. Nous remarquons également que le système a tendance à sur-compter, mais la précision globale peut encore être améliorée par l'introduction d'un facteur de correction. Cependant, les faux positifs sont davantage tolérés que les faux négatifs dans la plupart des applications. Le Tableau 2 compare la vitesse d'exécution de notre approche par rapport à d'autres approches (FPS : nombre d'images par seconde). Nous remarquons que notre approche est la plus adaptée pour les applications temps réel.

4 Conclusion

Dans cet article, nous avons présenté une nouvelle approche permettant de compter le nombre d'individus franchissant une ligne de comptage. Les principales étapes de notre méthodologie portent sur l'extraction des blobs franchissant cette ligne et l'estimation du nombre d'individus associés à ces blobs. L'approche proposée comporte trois originalités essentielles : (i) elle évite la détection des personnes statiques se trouvant sur la ligne de comptage, (ii) elle a recours à un algorithme de détection en ligne des blobs, et (iii) elle s'adapte facilement à différentes configurations de caméra grâce à un modèle de régression linéaire pour chaque configuration. Notre système est aussi performant pour gérer plusieurs lignes de comptage de manière robuste et efficace. Dans nos travaux futurs, nous nous intéresserons à l'introduction

d'un facteur de correction et à la création d'une base de vidéos utilisable par la communauté pour évaluer les systèmes de comptage pour des scénarios complexes.

Remerciements : Ce travail a été réalisé dans le cadre du projet ANR *CANADA* (2007-2010, Analyse des comportements Anormaux : Alerte, Détection, Action) et le projet européen Miauce (2006-2009, FP6, Call 5, IST-2005-5-033715).

Références

- Antic, B., D. Letic, D. Culibrk, et V. Crnojevic (2009). K-means based segmentation for real-time zenithal people counting. In *International Conference on Image Processing (ICIP)*.
- Benabbas, Y., N. Ihaddadene, T. Yahiaoui, T. Urruty, et C. Djeraba (2010). Spatio-temporal optical flow analysis for people counting. In *IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*.
- Chan, A. B., Z. S. J. Liang, et N. Vasconcelos (2008). Privacy preserving crowd monitoring : Counting people without people models or tracking. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Chen, D.-Y., C.-W. Su, Y.-C. Zeng, S.-W. Sun, W.-R. Lai, et H.-Y. M. Liao (2009). An online people counting system for electronic advertising machines. In *International conference on Multimedia and Expo (ICME)*.
- Cong, Y., H. Gong, S.-C. Zhu, et Y. Tang (2009). Flow mosaicking : Real-time pedestrian counting without scene-specific learning. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Djeraba, C., A. Lablack, et Y. Benabbas (2010). *Multi-Modal User Interactions in Controlled Environments*. Springer-Verlag.
- Enzweiler, M. et D. M. Gavrilu (2009). Monocular pedestrian detection : Survey and experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 31(12), 2179–2195.
- Lucas, B. D. et T. Kanade (1981). An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artificial Intelligence (IJCAI)*.
- Velipasalar, S., Y. li Tian, et A. Hampapur (2006). Automatic counting of interacting people by using a single uncalibrated camera. In *International Conference on Multimedia and Expo (ICME)*.
- Yang, D. B., H. H. Gonzalez-Banos, et L. J. Guibas (2003). Counting people in crowds with a real-time network of simple image sensors. In *International Conference on Computer Vision (ICCV)*.
- Zhao, X., E. Delleandrea, et L. Chen (2009). A people counting system based on face detection and tracking in a video. In *International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*.

Summary

In this paper, we present a new approach to count the number of people that cross a counting line from monocular video images. The proposed approach accumulates image slices and estimates the optical flow on them. Then, it performs an online blob detection on these slices in order to extract the crossing persons. The number of persons associated to each blob is determined using a linear regression model applied to blob features which are the position, velocity, orientation and size. The real-time performance and the high counting accuracy of this approach are demonstrated using on several datasets captured using either a vertical overhead or an oblique mounted camera.