HAL open science

Proposition et test de descripteurs pour l'identification de personnes dans un réseau de caméras

Ahmed Derbel, Sylvie Treuillet, Bruno Emile, Yousra Ben Jemaa, Abdelmajid Ben Hamadou

To cite this version:

Ahmed Derbel, Sylvie Treuillet, Bruno Emile, Yousra Ben Jemaa, Abdelmajid Ben Hamadou. Proposition et test de descripteurs pour l'identification de personnes dans un réseau de caméras. GRETSI 2011, Sep 2011, Bordeaux, France. pp.ID61. hal-00648495

HAL Id: hal-00648495

https://hal.science/hal-00648495v1

Submitted on 5 Dec 2011

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.

Proposition et test de descripteurs pour l'identification de personnes dans un réseau de caméras

AHMED DERBEL[1,2], SYLVIE TREUILLET[1], BRUNO EMILE[1], YOUSRA BEN JEMAA[3], ABDELMAJID BEN HAMADOU[2]

[1]Laboratoire PRISME
12 rue de Blois, F-45067 Orléans cedex2

[2]Laboratoire MIRACL
route de Tunis km 10, PB-3021-Sfax, Tunisie

[3]Unité Signaux et Systèmes
ENIT, le belvédère 1002, Tunis

ahmed.derbel@gmail.com, sylvie.treuillet@univ-orleans.fr,Bruno.Emile@univ-orleans.fr
yousra.benjemaa@enis.rnu.tn, abdelmajid.benhamadou@isims.rnu.tn

Résumé – Ce papier propose un descripteur basé sur les histogrammes régionaux fusionnant des caractéristiques intra et inter personnes pour faire l'indentification de piétons. Une étude comparative entre le descripteur proposé et les principaux vecteurs caractéristiques de la littérature montre que le vecteur proposé est capable d'identifier les personnes en multi-caméras face à la nature très articulée du corps humain, à la variation de la pose, à la variation des conditions d'éclairage, à la variation d'échelle, à la variation de points de vue et face aux occultations partielles.

Abstract – This paper proposes a feature based on regional histogram merging characteristics intra and inter persons to identify pedestrians. A comparative study between the proposed descriptor and the main feature vectors of the literature shows clearly that the proposed vector is able to identify people in multi-cameras despite the highly articulated human body, pose variation, lighting conditions, large scale and point of view variations, and some partial occlusions.

1 Introduction

L'identification et le suivi de personnes dans un réseau de caméras intéressent de nombreux domaines comme la vidéo surveillance, l'assistance à l'autonomie ou à la navigation pour les personnes âgées. La principale difficulté est de trouver des descripteurs sélectifs, caractérisant une personne parmi d'autres, tout en offrant une grande robustesse face aux fortes variations d'apparence induites par la nature très articulée du corps humain, et les changements d'éclairage et de points de vue selon les caméras.
Dans ce cadre, plusieurs méthodes ont été proposées dans la littérature essentiellement basées sur des descripteurs d'apparence, de formes et de couleur (histogramme, spatiogrammes, matrice de co-occurrence, histogrammes de gradients orientés), qui peuvent être combinées à des caractéristiques de mouvement ou à un suivi prédictif (filtres de Kalman [1], ou particulaire [2]).
La technique de comparaison adoptée analyse l'évolution des descripteurs entre deux images temporellement proches par une mesure de distance. Elle ne nécessite ni une phase d'apprentissage hors ligne, comme lors de l'utilisation de classifieurs [3], ni une réduction de dimensionnalité comme dans la mise en correspondance [4].
Dans cet article, nous proposons une variante des histogrammes régionaux couleurs pour représenter une personne qui combine des mesures d'homogénéité intra et inter personnes.
Cet article est structuré comme suit. La section suivante présente les différents vecteurs caractéristiques comparés. Les résultats trouvés sur différentes bases de piétons constituent l'objet de la troisième section. La dernière section présente la conclusion et les perspectives de ce travail.

2 Vecteurs caractéristiques

Dans cette section, nous présentons les descripteurs les plus usités ainsi que le descripteur proposé. Une phase de prétraitement est appliquée avant de calculer le vecteur caractéristique de chaque personne.

2.1 Prétraitement

La phase de prétraitement consiste à extraire les piétons supposés comme les seuls objets mobiles dans la scène.
Le processus suit la démarche suivante : 1) éliminer l'arrière plan [5]; 2) appliquer un filtrage morphologique; 3) localiser les personnes; 4) normaliser spatialement les images des per-

sonnes par une interpolation bicubique [100x35]; 5) normaliser les couleurs par Greyworld normalisation [6] (division de chaque valeur de pixel par la moyenne de la composante).

2.2 Descripteurs existants

L'objectif de cette section est de présenter les descripteurs les plus utilisés dans la littérature.

2.2.1 Histogrammes de couleurs

Les histogrammes caractérisent la distribution des couleurs dans l'imagette contenant un humain. Afin de réduire fortement la dimension du vecteur, une phase de quantification transforme les intensités des 3 canaux RVB codés sur 8 bits en 8 classes. L'histogramme de couleur est calculé sur l'image quantifiée ce qui donne un vecteur caractéristique de taille 8.

2.2.2 Spatiogrammes

Les spatiogrammes intègrent une information spatiale qui augmente significativement la robustesse des histogrammes de couleur [4]. Nous avons choisi de tester le spatiogramme proposé par C.O Conaire pour sa performance [7]. En plus d'un histogramme de couleur calculé comme expliqué en 2.2.1, on ajoute à chaque classe de l'histogramme sa distribution spatiale dans l'image quantifiée. Cette information spatiale est représentée par une gaussienne en fonction de la moyenne et la covariance des intensités observées dans l'image quantifiée. La variation entre chaque classe de l'histogramme va être pondérée par son homologue spatial, ce qui permet de mieux discriminer les personnes.

2.2.3 Les matrices de co-occurrences

Les matrices de co-occurrences caractérisent la texture en comptabilisant le nombre de fois où deux pixels de même intensité sont voisins dans l'image quantifiée selon une distance (généralement calculée en termes de pixels) et une direction donnée. Pour représenter une personne, nous avons choisi de travailler sur un voisinage d'ordre 1 (distance=1) et de sommer les matrices d'adjacence calculées selon 4 directions (0, 45, 90 et 135 degrés). Le vecteur caractéristique généré est de taille 8x8 pour chaque personne.

2.2.4 Histogrammes de gradients orientés (HGO)

Les histogrammes des gradients orientés (HGO) décrivent le contenu d'une image par la distribution des contours. Un filtre gaussien est préalablement appliqué sur l'image pour réduire le bruit avant l'extraction des contours [8]. L'image est ensuite divisée en 16 cellules non chevauchantes sur lesquelles on calcule un histogramme de l'orientation du contour selon 5 directions (horizontale, verticale, deux diagonales et non-uniforme). Le vecteur caractéristique généré est de taille 16*5 paramètres.

2.3 Descripteur proposé

L'idée de base de notre descripteur est de fusionner l'information intra et inter classes pour assurer une bonne signature

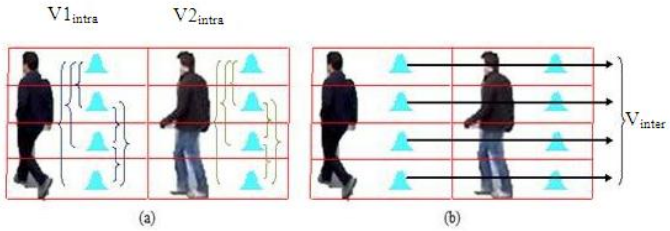

FIG. 1: *Descripteur proposé.*

des personnes. La première étape consiste à diviser horizontalement l'image de la personne en 4 blocs pour calculer l'histogramme de couleur régional de chaque bloc (Figure1). Cette division permet de prendre en considération la nature articulée et les différences d'apparence de couleurs entre les parties du corps humain. Pour chaque personne, deux vecteurs caractéristiques sont estimés. Le premier vecteur de taille 6 contient les distances de Battacharaya [9] entre les histogrammes de couleurs des 4 bandes : il caractérise la ressemblance intra classe (Figure1.a). Un second vecteur de taille 4 contient les distances de Battacharaya entre deux histogrammes de couleurs de la même bande pour deux personnes différentes (Figure1.b): il caractérise la ressemblance inter classes. La similarité moyenne est la mesure de ressemblance calculée entre deux piétons selon l'équation (1).

$$S_{moy_{m,i|n,j}} = \frac{1}{T}\sum_{t=1}^{T} S_{m,i,t|n,j,t+1} \tag{1}$$

Avec :

$$T = min(nb_{m,i}, nb_{n,j}) - 1$$

$$S_{m,i,t|n,j,t+1} = [1 - D_{m,i,t|n,j,t+1}] \times 100$$

$$D_{m,i,t|n,j,t+1} = w1 \times std(Vinter_{m,i,t|n,j,t+1}) + w2 \times std(Vintra_{m,i,t} - Vintra_{n,j,t+1})$$

$$Vinter_{m,i,t|n,j,t+1}(b) = d_B(Hc_{m,i,t,b}, Hc_{n,j,t+1,b})$$

$$Vintra_{m,i,t}(b,f) = d_B(Hc_{m,i,t,b}, Hc_{m,i,t,f}) \text{ où b} \neq \text{f}$$

Où m et n représentent l'identité des personnes, i et j représentent les différentes caméras utilisées, t représente l'ordre de l'image de la personne dans la séquence, b et f représentent le numéro du bloc dans la personne, S et D représentent respectivement la similitude et la différence entre deux images de personnes, T représente le nombre d'appariements effectués pour comparer deux séquences de piétons, nb est le nombre d'images dans une séquence, std est l'écart type d'un vecteur caractéristique qui représente dans notre cas une mesure de différence, Vintra et Vinter représentent respectivement les vecteurs caractéristiques intra et inter personnes, Hc est un histogra-

mme de couleur régional et d_B est la distance de Bhattacharaya. w1 et w2 sont les poids attribués respectivement l'information inter et intra personnes (w1=w2=0.5).

3 Expérimentation

3.1 Suivi de personnes

Pour évaluer la robustesse des vecteurs caractéristiques introduits en terme de suivi de personnes, une première étude comparative est réalisée sur deux séquences filmées chacune par trois caméras différentes. Ces séquences intègrent tous les problèmes existants dans le domaine de suivi des personnes à savoir la variation de la pose, la variation des conditions d'éclairage, la variation d'échelle, la variation de points de vue et les occultations partielles. La première séquence est extraite de la base PETS2009 [10] contenant 606 images de 4 personnes et la deuxième est extraite de la base [11] contenant 900 images de 3 personnes. La méthode de comparaison proposée entre les vecteurs caractéristiques présentés dans cet article est basée sur le calcul de la moyenne de la similitude entre des images successives des différentes personnes. Pour les histogrammes de couleurs, les spatiogrammes, les histogrammes de gradients orientés et les matrices de co-occurrences la similitude entre deux vecteurs caractéristiques est égale à la distance de Bhattacharaya. Pour notre descripteur, la similitude est calculée par l'équation (1) sur des vecteurs caractéristiques contenant des distances de Bhattacharaya. La figure(2) représente pour les deux bases la similitude moyenne pour la même personne et la similitude moyenne pour des personnes différentes. Le tableau(1) présente le nombre de fausses alertes (c'est le nombre de fois où la similitude entre la même personne est inférieure à la similitude entre deux personnes différentes) pour les deux bases de piétons.

Les résultats montrent que les histogrammes de couleurs et les spatiogrammes permettent de suivre les personnes dans un réseau de caméras. Ces descripteurs peuvent contourner la grande variation de la pose, les occultations partielles générées par des erreurs de classification des pixels avant plan et des chevauchements entre personnes. La supériorité des spatiogrammes par rapport aux histogrammes de couleurs est relative à l'intégration de l'information spatiale capable de rapprocher d'avantage les personnes ayant des poses relativement proches. Ces deux descripteurs restent sensibles à la grande variation de luminosité inter-caméras qui affectent les couleurs dans les silhouettes des personnes. Ces résultats montrent aussi que les HGO et les matrices de co-occurrences ne sont pas adaptés pour suivre les personnes en multi-caméras à cause de leurs grandes sensibilités à la variation de la pose. On peut remarquer que les matrices de co-occurrences sont plus performantes que les HGO vu que la texture subit moins de variations que les contours en multi-caméras.

L'histrogramme régional introduit dans cet article présente un descripteur performant héritant le pouvoir descriminant des histrogrammes de couleurs avec l'intégration de l'information intra classe qui limite l'impact négatif de la variation des conditions d'éclairage inter-caméras.

TAB. 1: *le nombre de fausses alertes pour les deux bases (PETS 2009, CVLAB)*

	Nombre de fausses alertes	
	PETS2009	CVLAB
Histogramme de couleur	2	2
spatiogramme	2	2
HGO	11	4
Matrice de co-occurrence	4	2
Descripteur proposé	0	1

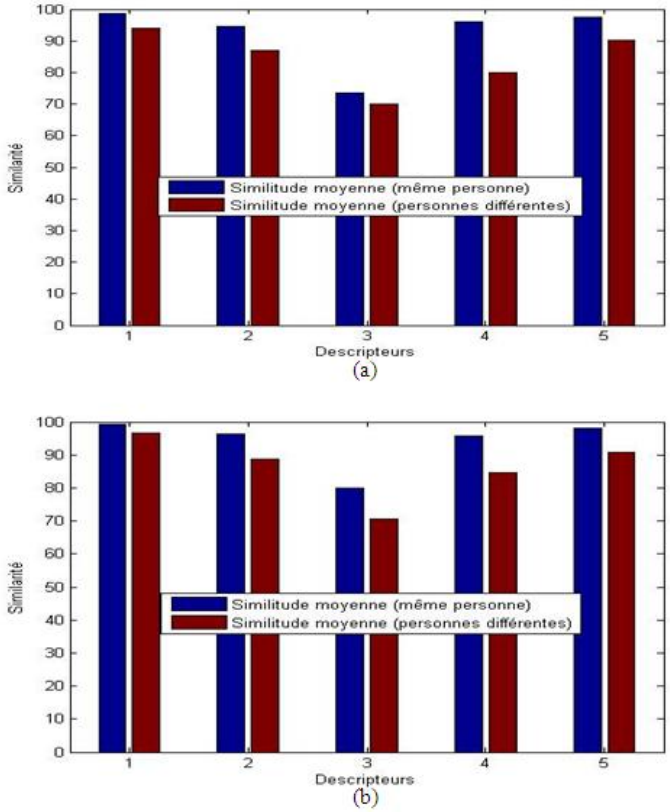

FIG. 2: *(a) La similitude moyenne pour la base PETS2009, (b) La similitude moyenne pour la base CVLAB. (1 : histogramme de couleur, 2 spatiogramme, 3 histogramme de gradients orienté, 4 matrice de co-occurrences et 5 est le descripteur proposé).*

3.2 Identification de personnes

L'objectif de cette deuxième expérimentation est d'évaluer la robustesse des vecteurs caractéristiques présentés dans cet article en terme d'identification de personnes sur la base VIPeR [12]. Cette base contient 632 paires d'images de piétons (2 images par personne). Chaque paire d'images a été prise à par-

tir d'un point de vue arbitraire dans des conditions d'éclairage variables. Dans cette expérience, nous avons utilisé la mesure CMC "cumulative matching characteristic" sur la moitié de cette base (316 piétons). Le (CMC) est une mesure de la précision d'un système biométrique. La courbe CMC présente le taux de bonne identification si on ne considère que les n premiers résultats retournés par le système (n représente le rang). Nous avons comparé les résultats des vecteurs caractéristiques déjà présentés avec l'approche [13]. Cette approche [13] est basée sur une cascade de descripteurs (couleur et texture) pour faire l'identification de personnes. Les résultats de cette expérience sont illustrés dans la figure(3).

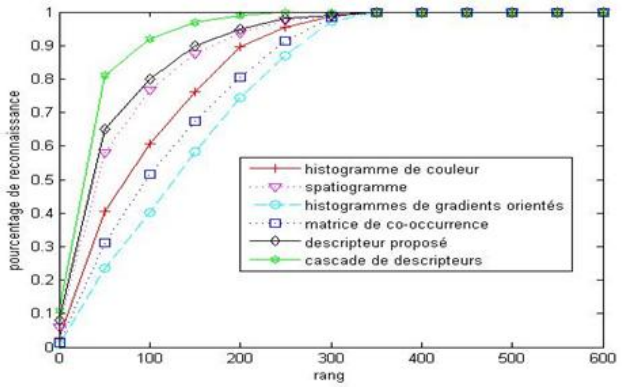

FIG. 3: *Courbe CMC pour les différents descripteurs.*

La figure(3) montre la robustesse du vecteur caractéristique proposé face aux histogrammes de couleur, aux spatiogrammes, aux matrices de co-occurrences et aux histogrammes de gradients orientés. La figure(3) montre aussi la supériorité de l'approche [13] par rapport à notre vecteur caractéristique vue qu'elle fusionne des descripteurs de couleur et de texture complémentaires en cascade.

4 Concusion et perspectives

Après cette étude comparative, on peut conclure que la fusion des caractéristiques intra et inter personnes présente un descripteur performant assurant une bonne identification de personnes en multi-caméras. L'avantage du descripteur proposé est expliqué d'une part par l'utilisation des caractéristiques intra personnes qui sont moins sensibles à la grande variation des couleurs causée par le changement des conditions d'éclairage entre caméras.
D'autre part, il est expliqué par l'intégration des caractéristiques inter personnes qui sont robustes face aux changements de poses et aux occultations partielles.
Comme perspective à ce travail, nous allons essayer de fusionner le descripteur proposé avec d'autres vecteur caractéristiques ce qui augmentera d'avantage le taux de bonne identification des personnes en multi-caméras.

Références

[1] E. Monari, J. Maerker et K. Kroschel. A Robust and Efficient Approach for Human Tracking in Multi-camera Systems. AVSS, pp.134–139, 2009.

[2] C. Yang, R. Duraiswami et L. Davis. Fast multiple object tracking via a hierarchical particle filter. International Conference on Computer Vision, pp. 212–219, 2005.

[3] C. Nakajima, M. Pontil, M. Heisel et T. Poggio. Full body person recognition system. Pattern Recognition. 36(9), pp. 1997–2006, 2003.

[4] D.N. Truong Cong, L. Khoudour et C. Achard. Approche spectrale et descripteur couleur-position statistique pour la réidentification de personnes à travers un réseau de caméras. GRETSI, 2009.

[5] A. Elgammal, D. Hawood et L. Davis. Non-parametric model for background subtraction. In: 6th European Conference on Computer Vision FRAME-RATE Workshop, 2000.

[6] G. Buchsbaum. A spatial processor model for object color perception. Journal of the Franklin Institute. 310 (1), pp. 1–26, 1980.

[7] C. Conaire, N. O'Connor et A. Smeaton. An improved spatiogram similarity measure for robust object localization. In: 32th IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 1069–1072. Hawaii, 2007.

[8] C. Yang, R. Duraiswami et L. Davis. Fast multiple object tracking via a hierarchical particle filter. In: 10th International Conference on Computer Vision, pp. 212–219. Beijing, 2005.

[9] A. Bhattacharyya. On a measure of divergence between two statistical populations defined by probability distributions. Bulletin of the Calcutta Mathematical Society. 35, pp. 99–109, 1943.

[10] Eleventh IEEE Internationa Workshop on Performance Evaluation of Tracking and Surveillance, ftp://ftp.cs.rdg.ac.uk/pub/PETS2009/.

[11] Computer Vision Laboratory (CVLAB), http://cvlab.epfl.ch/data/pom/.

[12] VIPeR: Viewpoint Invariant Pedestrian Recognition, http://vision.soe.ucsc.edu/node/178

[13] D. Gray et H. Tao. Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features. In Proc. European Conference on Computer Vision (ECCV), 2008.