Information Fusion xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Information fusion to detect and classify pedestrians using invariant features

Ana Pérez Grassi [a,*], Vadim Frolov [b], Fernando Puente León [b]

[a] Institute for Measurement Systems and Sensor Technology, Technische Universität München, Theresienestr. 90/N5, 80333 Munich, Germany
[b] Institut für Industrielle Informationstechnik, Karlsruhe Institute of Technology, Hertzstr. 16/Geb. 06.35, D-76187 Karlsruhe, Germany

ARTICLE INFO

Article history:
Received 4 February 2009
Received in revised form 14 April 2010
Accepted 15 June 2010
Available online xxxx

Keywords:
Pedestrian detection
Invariant
Classification
Data fusion
Infrared
Lidar
Spatio-temporal fusion
Support vector machines

ABSTRACT

A novel approach to detect pedestrians and to classify them according to their moving direction and relative speed is presented in this paper. This work focuses on the recognition of pedestrian lateral movements, namely: walking and running in both directions, as well as no movement. The perception of the environment is performed through a lidar sensor and an infrared camera. Both sensor signals are fused to determine regions of interest in the video data. The classification of these regions is based on the extraction of 2D translation invariant features, which are constructed by integrating over the transformation group. Special polynomial kernel functions are defined in order to obtain a good separability between the classes. Support vector machine classifiers are used in different configurations to classify the invariants. The proposed approach was evaluated offline considering fixed sensors. Results obtained based on real traffic scenes demonstrate very good detection and classification rates.

1. Introduction

The ability to detect and classify human beings is fundamental in building intelligent systems for applications like visual surveillance, robotics, autonomous vehicles, and driver assistance systems. In recent years, driver assistance systems have particularly attracted attention of researchers all over the world. The principal motivation is to minimize the number of deaths in car accidents. Only in Europe, about 40,000 victims of car accidents have been reported during 2007 [1]. Passive safety systems (like air bags and belts) can decrease the number of fatal accidents, while active safety systems can reduce the number of accidents. Additionally, passive safety systems are designed to protect the car drivers and passengers, whereas active systems can extend this protection to other traffic participants. Thus, one of the most important tasks of active safety systems for urban environments is the recognition of pedestrians.

Pedestrian recognition is a challenging task due to the variability of their appearances and poses. Moreover, the background of a traffic scene is incredibly unpredictable, which makes the pedestrian segmentation especially difficult. Various kinds of vehicle-based sensors are used to solve this task. Commonly used sensors are passive imaging sensors using visible light and infrared (IR) radiation, as well as active time-of-flight sensors, such as radar and lidar scanners. Imaging sensors are widely used because of their high lateral resolution and low cost, but extracting information from them involves substantial amount of processing. Furthermore, these sensors are very sensitive to the environment illumination and weather conditions. Time-of-flight sensors provide information about objects distances, but they do not deliver enough data to perform a complex classification. These two types of sensors complement each another, and their fusion is expected to present better results than single-sensor systems [2].

This paper presents a new method of detecting pedestrians and classifying them according to their movement patterns. The approach is based on the signals of an IR camera and a lidar scanner as well as the extraction of invariant features. Recent works to detect pedestrians using infrared cameras can be found in [2–4]. All of these works are based on the fusion of several imaging sensors data. Stereo-infrared is used in [3], while [2] and [4] fuse visual and IR data. Approaches that use time-of-flight sensors are described in [5–7]. The fusion of image and time-of-flight sensors has been studied in [8–11]. All these works are limited to the detection of pedestrians. Some approaches to classify human behavior were developed in [12–14]. In [12], human behavior estimations are made in the context of video surveillance, while in [13] motion patterns are used to avoid vehicle-to-pedestrian collisions.

To recognize pedestrians, a representative set of features has to be extracted from the raw data. State-of-the-art techniques use features based on shape, motion or depth information. Some of the

* Corresponding author.
E-mail addresses: a.perez@tum.de (A. Pérez Grassi), vadim.frolov@kit.edu (V. Frolov), puente@kit.edu (F. Puente León).

1566-2535/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.inffus.2010.06.002

features used for shape-based detection are size and aspect ratio of bounding boxes [3], Haar wavelets [15], Haar-like wavelets [11], pose-specific linear (PCA) features [16], active contours [17], invariant features [18], scale-invariant DoG features [19,20], intensity gradients [21] and their histograms [22,23]. Typical motion features include signal moments [24,25], symmetry characteristics of the legs [14,26], gait patterns [24,27] and motion patterns [28]. A combination of shape and motion information is presented by Viola et al. [29] in order to detect pedestrians in low-resolution images under difficult weather conditions. Global appearance changes caused by pedestrian articulations and different viewpoints are considered in [30]. Tracking techniques are often used in pedestrian detection as well. In [31], the speed and the size of segmented objects are used to create hypotheses, which are verified with a vision system. The fusion of clustered objects in four horizontal laser planes is presented in [5]. In [6], objects are tracked, and their shapes are recursively estimated in order to detect pedestrians.

The features extracted from the raw data must be classified. Various types of classifiers are used to distinguish pedestrians from other objects. Some of the commonly used classifiers are Support Vector Machines (SVMs) [2,15,16,22,27], AdaBoost [29] and various types of neural networks.

The presented approach is based on the extraction and classification of invariant features resulting from the fusion data of an IR camera and a lidar scanner. The principal objective of this work is to obtain more relevant information about the traffic participants than that achieved through a simple detection. The most important contribution of this paper is that pedestrians are not only detected, but also classified according to their movements, which is achieved without using any tracking technique or movement analysis. Both detection and classification are performed through different fusion levels. Additionally, the proposed invariants are compared to the well-known Histograms of Oriented Gradient (HOG) features [32]. As shown later, classification results based on the presented features outperform those based on HOG. The improvement over HOG is particularly significant in the classification of the pedestrian movements. Although the presented results are generated offline, the obtained classification rates show that the proposed method is a promising approach, as it can be easily adapted to real application conditions.

This paper is organized according to the different data fusion levels considered. Section 2 describes the first data fusion, which is performed at a signal level on the IR and lidar data in order to determine regions of interest (ROI) in the IR pictures. Section 3 presents the extraction of invariant features from the ROIs, which results in a spatio-temporal fusion. The latest fusion level, analyzed in Section 4, is performed by the classification of the features through different SVM configurations. Finally, Section 5 presents detection and classification results and compares the proposed method to the HOG approach.

2. Extraction of regions of interest

2.1. Sensor signals: lidar and infrared camera

The presented approach is based on two sensors: an infrared camera and a lidar scanner. The infrared camera yields information about shape and temperature of the objects in the traffic scene in form of a digital video. The video, denoted by $\mathcal{G}$, can be defined as a sequence of frames in the time:

$$\mathcal{G} = \langle g^k(\mathbf{m}), k \in \{0, \ldots\}\rangle. \tag{1}$$

Each video frame has a size of $M \times N$ pixels and is represented by:

$$g^k(\mathbf{m}) := g(\mathbf{m}, k\Delta t), \tag{2}$$

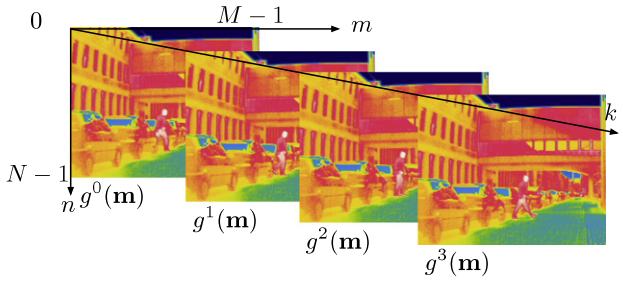


Fig. 1. Camera output: digital video $\mathcal{G}$ for $k \in \{0, 1, 2, 3\}$.

where $\mathbf{m} = (m\Delta x, n\Delta y)^T$ with $m \in \{0, \ldots, M-1\}$ and $n \in \{0, \ldots, N-1\}$. The recording speed is one frame per Δt seconds. The variables Δx, Δy, M, N and Δt are setup parameters of the camera. Fig. 1 schematizes the definition of $\mathcal{G}$.

The used lidar scanner performs a one-line scan of the scene with an angular resolution of $\Delta\varphi$ and an aperture angle of 180°. The scene is completely scanned each Δt seconds. Thus, the lidar signal $\mathcal{D}$ can also be described as a sequence in time:

$$\mathcal{D} = \langle d_w^k, k \in \{0, \ldots\}\rangle, \tag{3}$$

where $d_w^k := d(w\Delta\varphi, k\Delta t)$ and $w \in \{0, 1, \ldots, W = 180°/\Delta\varphi\}$. The signal d_w^k gives the distances to the objects in the scene, whose positions at the point in time $k\Delta t$ coincide with the scanning angles $w\Delta\varphi$.

2.2. Preprocessing of the infrared video

The intensity values of the infrared video $\mathcal{G}$ depend on the temperature of the imaged objects: the warmer an object is, the brighter it appears on the IR pictures. Because the temperature of the human body is mostly higher than that of its surroundings, people appear as bright objects in IR pictures. However, because of the isolating properties of some clothes, the body of a person can seldom be imaged as a whole warm object. Generally, only hands and heads appear clearly brighter than the surroundings [33,34]. Taking this into account, all warm objects on an IR picture can be segmented by a simple threshold. In this way, the information of the complete scene is reduced only to those objects that, due to their temperature, could potentially be part of a human body. If the intensity value ξ represents the minimal possible temperature associated with a human body, then the result of the threshold for a frame $g^k(\mathbf{m}) \in \mathcal{G}$ can be defined as follows:

$$\breve{g}^k(\mathbf{m}') = \begin{cases} +1 & g^k(\mathbf{m}') \geqslant \xi \\ -1 & \text{elsewhere.} \end{cases} \tag{4}$$

The positive values of $\breve{g}^k(\mathbf{m})$ are clustered according to their proximity by means of morphological operations. The resulting hot spots are labeled by an index $q_k \in \{1, \ldots, Q_k\}$ and denoted by h_k^q. Each of the extracted Q_k hot spots is considered a potential human head (See Fig. 2B). If occlusion is not considered, the resulting hot spots h_k^q include all human heads in the scene. In other words, this procedure step produces no false negatives. However, in urban scenes,

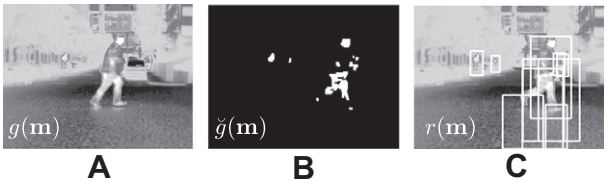


Fig. 2. Resulting hot spots h_k^q from an infrared image. A: infrared image showing five pedestrians. B: Ten hot spots h_k^q resulting from applying a threshold. C: Ten ROIs determined by the hot spots and the lidar information (all pedestrians were extracted).

Please cite this article in press as: A. Pérez Grassi et al., Information fusion to detect and classify pedestrians using invariant features, Informat. Fusion (2010), doi:10.1016/j.inffus.2010.06.002

there are many objects, like car parts, traffic signals, building windows, etc., that also generate hot spots in $\breve{g}^k(\mathbf{m})$. On average, around 80% of the extracted hot spots are related to non-pedestrian objects, i.e., to false positives. This set of non-pedestrian objects should be discarded by the detection process described later.

2.3. Preprocessing of lidar data

For each value of k, the lidar values d_w^k are also grouped according to their proximity. The used clustering algorithm is presented in [35]. It is considered that each resulting group forms part of the same object in the real world. These lidar objects are labeled by an index $p_k \in \{1, \ldots, P_k\}$ and denoted by o_k^p. The distance between the sensor and the center of gravity of each object o_k^p is given by $d_w(o_k^p)$ [18,35].

2.4. Lidar and infrared data correlation

As the camera and lidar speeds have been defined to be equal, both sensors are synchronized. Now, the signals must be registered in order to associate each hot spot h_k^q of the IR signal with a lidar object o_k^p. To obtain a correlation between the infrared and lidar data, both sensors are vertically aligned with the camera over the lidar sensor (see Fig. 3). In this way, the scanning projection of the lidar sensor coincides with the horizontal projection of the camera. An equivalence relation can be established between the coordinates $m\Delta x$ and $w\Delta\varphi$ for a certain point in time $k\Delta t$ as follows:

$$m\Delta x = \frac{M-1}{2(\cos(w_{\max}\varphi))}\cos(w\Delta\varphi) + \frac{M-1}{2}, \tag{5}$$

where w is now defined only in the range of interest, which coincides with the camera's field of view: $w \in \{w_{\min}, \ldots, w_{\max}\}$ with $\cos(w_{\max}) = -\cos(w_{\min})$. The described equivalency between the coordinates of both sensors is denoted by $m \equiv w$. A hot spot h_k^q and a lidar object o_k^p belong to the same object in the real world, if their respective coordinates m and w are equivalent. Now, a distance $d_m(h_k^q)$ can be established to each hot spot, where $d_m(h_k^q) = d_w(o_k^p)$ for $m \equiv w$. Fig. 4 illustrates the correspondence between hot spots h_k^q and lidar objects o_k^p.

2.5. ROI

As already mentioned, each hot spot h_k^q is potentially considered to be a human head. Starting from the position of each h_k^q, the complete hypothetical human body must be extracted from $g^k(\mathbf{m})$. A generic pedestrian size of $2{,}m \times 1{,}m$ is defined in world coordinates. Now, this real world size must be brought to image coordinates for each hot spot h_k^q. This can be performed using the corresponding distances $d_m(h_k^q)$ and a pinhole camera projection model. Then, for each potential human head h_k^q, a region of $M_q \times N_q$ pixels is defined, which may contain the complete body. These ROIs are denoted by $r_q^k(\mathbf{m}_q)$, where $\mathbf{m}_q = (m_q\Delta x, n_q\Delta y)^{\mathrm{T}}$ with $m_q \in \{0, \ldots, M_q - 1\}$ and $n_q \in \{0, \ldots, N_q - 1\}$. The size of $r_q^k(\mathbf{m}_q)$ is a function on q, i.e., a function on the distance $d_m(h_k^q)$ (See Fig. 2C).

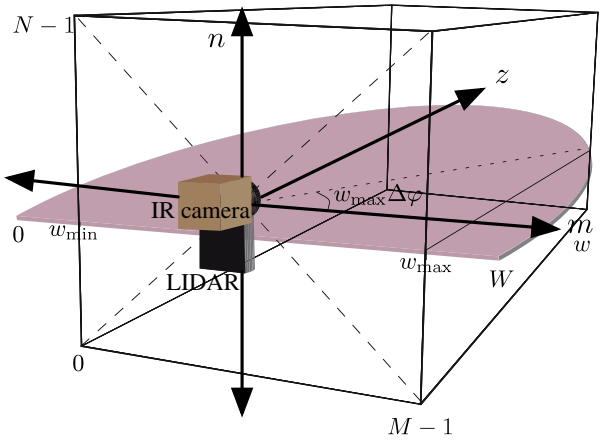


Fig. 3. Sensor platform setup.

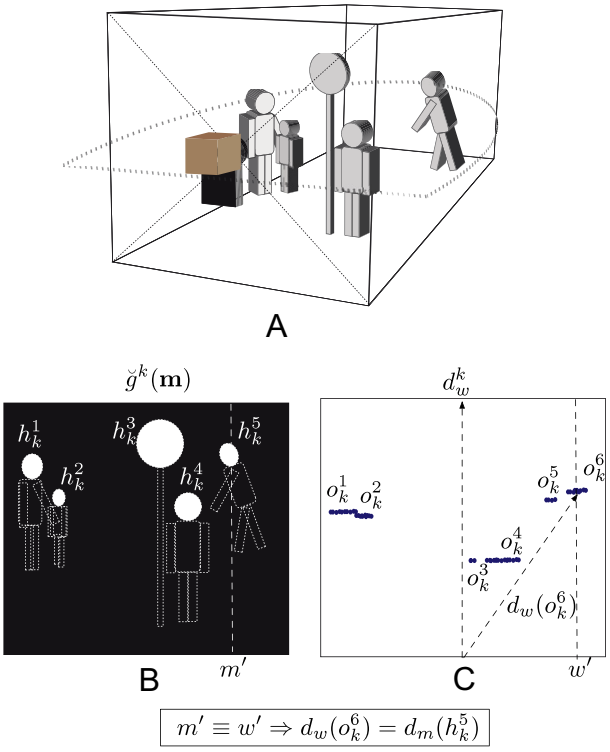


Fig. 4. A: Simulation of a traffic scene with four pedestrians. B: Hot spots h_k^q from $\breve{g}^k(\mathbf{m})$, with $Q_k = 5$ (one of the extracted hot spot does not correspond with a pedestrian). C: lidar objects o_k^p, with $P_k = 6$ (one pedestrian was clustered incorrectly as two different objects).

The presented approach constitutes a system with memory. This means that the classification results at a time point $k = a$ depend not only on the actual signal values $g^a(\mathbf{m})$ and d_w^a, but also on the previous values of $g^k(\mathbf{m})$ for $k < a$. To consider past IR information, each ROI $r_q^k(\mathbf{m}_q)$ is extracted not only from $g^k(\mathbf{m})$, but also from K frames in the past. Then, the final ROI can be defined as a sequence:

$$\mathcal{R}_k^q = \langle r_q^{k'}(\mathbf{m}_q), k' \in \{k, \ldots, k-K\}\rangle, \tag{6}$$

where, as mentioned before, K defines the number of frames observed in the past. Fig. 5 illustrates the extraction of two ROIs.

As mentioned $M_q \times N_q$ depends on the distances $d_m(h_k^q)$. This dependency must be avoided in order to suppress scale transformations. For this purpose, the sizes of all ROIs $\mathcal{R}_k^q$ are scaled through interpolation to a normalized size $\check{M} \times \check{N}$. The resulting normalized ROIs are denoted by:

$$\check{\mathcal{R}}_k^q = \langle \check{r}_q^{k'}(\check{\mathbf{m}}), k' \in \{k, \ldots, k-K\}\rangle, \tag{7}$$

where $\check{\mathbf{m}} = (\check{m}\Delta x, \check{n}\Delta y)^{\mathrm{T}}$ with $\check{m} \in \{0, \ldots, \check{M} - 1\}$ and $\check{n} \in \{0, \ldots, \check{N} - 1\}$.

This section has described the extraction of ROIs from the IR video signal $\mathcal{G}$. The complete process results in a data in-data out (DAI-DAO) fusion [36]. Fig. 6 illustrates the ROI extraction method.

Please cite this article in press as: A. Pérez Grassi et al., Information fusion to detect and classify pedestrians using invariant features, Informat. Fusion (2010), doi:10.1016/j.inffus.2010.06.002

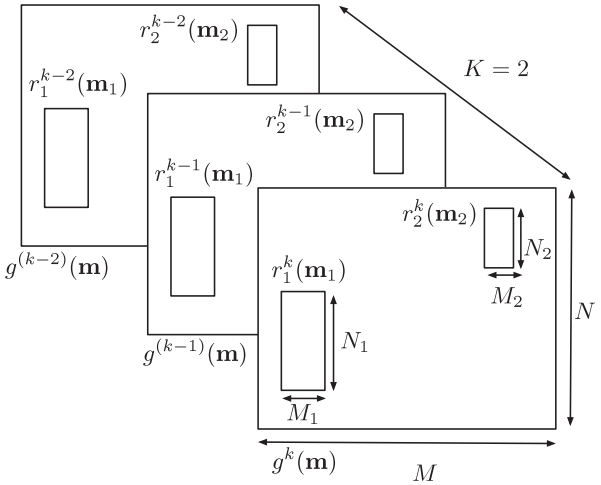


Fig. 5. Extraction of $\mathcal{R}_k^1$ and $\mathcal{R}_k^2$ for $K = 2$ and $d_m(h_k^1) < d_m(h_k^2)$.

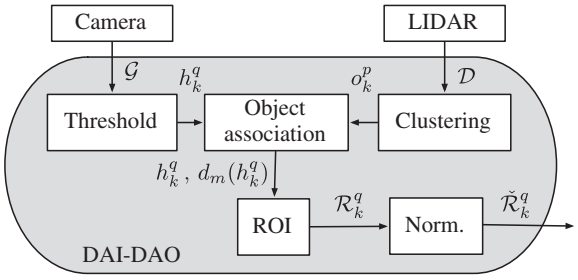


Fig. 6. Scheme of the ROI extraction process.

3. Extraction of invariants

The normalized ROIs $\check{\mathcal{R}}_k^q$ are the patterns that must be classified. This classification is based on invariant features extracted from each $\check{\mathcal{R}}_k^q$. In this section, an approach to construct these invariants by integration will be presented.

3.1. Invariants by integration

Objects in the real world can be affected by transformations, but that should not alter their classification. These transformations in the real world induce transformations in the pattern space. For a recognition task, different patterns are considered equivalent if they convey to each other through an induced transformation [37]. An induced transformation $\mathrm{T}(t)$ on a pattern $\check{\mathcal{R}}_k^q$ is defined as a bijective mapping [38]:

$$\mathrm{T} : (\check{\mathcal{R}}_k^q, t) \mapsto \mathrm{T}(t)\check{\mathcal{R}}_k^q \quad \forall t \in \mathcal{T}, \tag{8}$$

where $\mathcal{T}$ is the set of all transformation parameters t. The set of all transformations is denoted by $\mathrm{T}(\mathcal{T}) = \{\ \mathrm{T}(t) \mid t \in \mathcal{T}\}$. The transformation set $\mathrm{T}(\mathcal{T})$ defines an equivalence relation in the pattern space, where $\check{\mathcal{R}}_k^q \equiv \mathrm{T}(t)\check{\mathcal{R}}_k^q$ for all $t \in \mathcal{T}$ [38].

A feature $\tilde{f}^l$ is called invariant if, for a given transformation set $\mathrm{T}(\mathcal{T})$, it remains constant for all equivalent patterns:

$$\tilde{f}^l(\check{\mathcal{R}}_k^q) = \tilde{f}^l(\mathrm{T}(t)\check{\mathcal{R}}_k^q) \quad \forall t \in \mathcal{T}. \tag{9}$$

If the set $\mathrm{T}(\mathcal{T})$ forms a compact group, then an invariant $\tilde{f}^l(\check{\mathcal{R}}_k^q)$ can be constructed by integrating over this group [37–39]:

$$\tilde{f}^l(\check{\mathcal{R}}_k^q) = \frac{1}{|\mathrm{T}(\mathcal{T})|}\int_{\mathcal{T}} f^l(\mathrm{T}(t)\check{\mathcal{R}}_k^q)\mathrm{d}t, \tag{10}$$

where $f^l(\mathrm{T}(t)\check{\mathcal{R}}_k^q) := f(\ \mathrm{T}(t)\check{\mathcal{R}}_k^q, \mathbf{w}_l)$ is a real function of the transformed pattern and a parameter vector $\mathbf{w}_l$. This function is called kernel function. The factor $|\mathrm{T}(\mathcal{T})|$ normalizes the result with respect to the group volume.

As defined in Eq. (7), the ROI $\check{\mathcal{R}}_k^q$ is a discrete signal. If the transformation is discretized by defining $\mathcal{T} = \{t^0, \ldots, t^{(T-1)}\}$, then the integral in Eq. (10) can be replaced by a summation:

$$\tilde{f}^l(\check{\mathcal{R}}_k^q) = \frac{1}{|\mathrm{T}(\mathcal{T})|}\sum_{\mathcal{T}} f^l(\mathrm{T}(t)\check{\mathcal{R}}_k^q,). \tag{11}$$

The calculation of the transformed pattern $\mathrm{T}(t)\check{\mathcal{R}}_k^q$ for each value $t \in \mathcal{T}$ is computationally intensive. A more efficient solution is to perform the transformation $\mathrm{T}(t)$ not on the pattern but on the kernel function [37]:

$$\mathrm{T}(t)\{f^l(\check{\mathcal{R}}_k^q)\} = f^l(\mathrm{T}(t^{-1})\check{\mathcal{R}}_k^q) = f_t^l(\check{\mathcal{R}}_k^q), \tag{12}$$

where f_t^l denotes the transformed kernel function. Eq. (11) can be rewritten as follows:

$$\tilde{f}^l(\check{\mathcal{R}}_k^q) = \frac{1}{|\mathrm{T}(\mathcal{T})|}\sum_{\mathcal{T}} f_t^l(\check{\mathcal{R}}_k^q) \tag{13}$$

3.2. Transformation group

For the detection and classification of pedestrians, the 2D translation constitutes the transformation group of interest. In this case, the transformation parameter is given by the translation vector $\mathbf{t}$: $= (i\Delta x, j\Delta y)^{\mathrm{T}}$, where $i \in \{0, \ldots, \check{M} - 1\}$ and $j \in \{0, \ldots, \check{\mathrm{N}} - 1\}$. In order to obtain invariants by integration, the translation in the ROIs must be considered cyclical. This way, the transformation group becomes compact. The transformed normalized ROI can be defined as follows:

$$\mathrm{T}(\mathbf{t})\check{\mathcal{R}}_k^q = \left\langle \check{r}_q^{k'}(\check{\mathbf{m}} + \mathbf{t})k' \in \{k, \ldots, k - K\}\right\rangle \tag{14}$$

where, for ease of exposition, the modulo of M and N (necessary to consider the translation cyclical) has been suppressed from the notation. Finally, after introducing this transformation group in Eq. (11), the invariant feature can be written as follows:

$$\tilde{f}^l(\check{\mathcal{R}}_k^q) = \frac{1}{\check{M}\check{N}}\sum_{i=0}^{\check{M}-1}\sum_{j=0}^{\check{N}-1} f^l(\mathrm{T}(\mathbf{t})\check{\mathcal{R}}_k^q). \tag{15}$$

3.3. Kernel function

The kernel function should be constructed to extract from $\check{\mathcal{R}}_k^q$ all relevant information for its classification. For this approach, the parameter vector of the selected kernel function is given by:

$$\mathbf{w}_l := (U_l, V_l, s, h(u_l, v_l), e(s)), \tag{16}$$

where $U_l, V_l \in \mathbb{N}$ and $s \in \mathbb{Z}$. The fourth element of $\mathbf{w}_l$ is a function of the variables $u_l \in \{0, \ldots, U_l - 1\}$ and $v_l \in \{0, \ldots, V_l - 1\}$, where $h(u_l, v_l) \in \mathbb{N}_0$. The last parameter is a function of s defined as follows:

$$e(s) = \begin{cases} K & |s| \leqslant \eta \\ K - \zeta & |s| > \eta, \end{cases} \tag{17}$$

where $\eta, \zeta \in \mathbb{N}$.

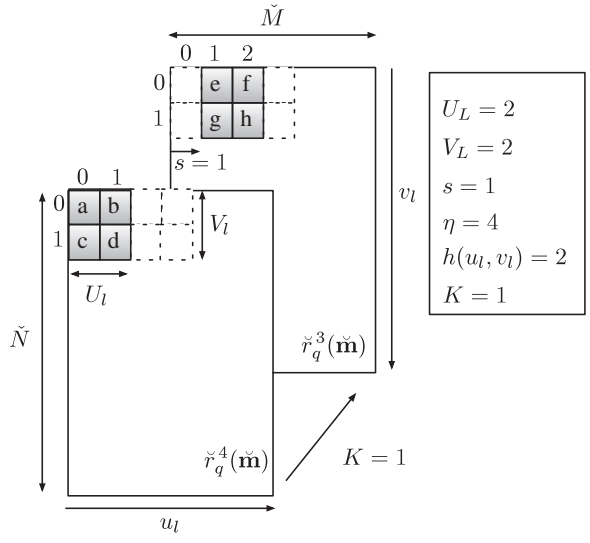
The selected kernel function f^l is a polynomial consisting of $e(s) + 1$ terms defined as follows:

$$f^l(\check{\mathcal{R}}_k^q) = \sum_{k'=k}^{k-e(s)}\prod_{v_l=0}^{V_l-1}\prod_{u_l=0}^{U_l-1}[\check{r}_q^{k'}(\mathbf{u}_l)]^{h(u_l,v_l)}, \tag{18}$$

where $\mathbf{u}_l = (u_l + s(k' - k), v_l)^{\mathrm{T}}$. Fig. 7 shows an example of a kernel function $f^l(\check{\mathcal{R}}_4^q)$ for $U_l = 2$, $V_l = 2$, $s = 1$, $h(u_l, v_l) = 2$ and $\eta = 4$, using $K = 1$ and letters to represent intensity values.

Please cite this article in press as: A. Pérez Grassi et al., Information fusion to detect and classify pedestrians using invariant features, Informat. Fusion (2010), doi:10.1016/j.inffus.2010.06.002

As expressed in Eq. (18) and as illustrated in Fig. 7, the summation over k' causes that each term of the polynomial collects information from a different frame. The expression $s(k-k')$ induces a horizontal displacement on the coordinates of each polynomial term. This displacement is inversely proportional to k', and its direction depends on the sign of s. The meaning of the parameter s can be understood by interpreting the extraction of a ROI as the observation of the scene through a window given by $\check{r}_q^{k'}(\check{\mathbf{m}})$. For $k'=k$, i.e., for the actual frame, the pedestrian will be approximately in the center of $\check{r}_q^{k'=k}(\check{\mathbf{m}})$. On the other hand, if the pedestrian is moving horizontally, he or she will appear in $\check{r}_q^{k'<k}(\check{\mathbf{m}})$ towards the left or the right side depending on his/her direction and speed (see Fig. 8). By adjusting s, the kernel function can be configured to "look for" pedestrians in the past. In this way, it is possible to construct kernel functions that are more sensitive to one movement than to others, which finally produces a bigger separability between the classes.



$$\begin{aligned} f^l\left(\check{\mathcal{R}}_4^q\right) &= [\check{r}_q^4(0,0)^{\mathrm{T}}]^2\cdot[\check{r}_q^4(1,0)^{\mathrm{T}}]^2\cdot[\check{r}_q^4(0,1)^{\mathrm{T}}]^2 \\ &\quad\cdot[\check{r}_q^4(1,1)^{\mathrm{T}}]^2+[\check{r}_q^3(1,0)^{\mathrm{T}}]^2\cdot[\check{r}_q^3(2,0)^{\mathrm{T}}]^2 \\ &\quad\cdot[\check{r}_q^3(1,1)^{\mathrm{T}}]^2\cdot[\check{r}_q^3(2,1)^{\mathrm{T}}]^2 \\ &= (\mathrm{abcd})^2+(\mathrm{efgh})^2 \end{aligned}$$

Fig. 7. Example of a kernel function.

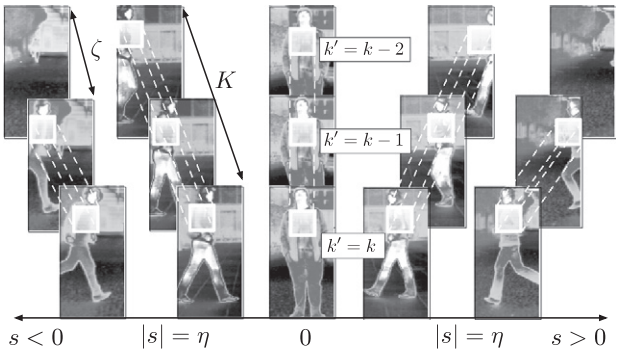


Fig. 8. Five different ROIs $\check{\mathcal{R}}_k^q$ with $K=3$. The five ROIs present pedestrians and have been aligned according to their classes: running to the left, walking to the left, no movement, walking to the right, running to the right. The small squares on each $\check{r}_q^{k'}(\check{\mathbf{m}})$ represent the calculation of a polynomial term of the kernel function.

Additionally, the faster a pedestrian moves, the shorter the time in which he appears in $\check{\mathcal{R}}_k^q$ will be. This fact is considered through the function $e(s)$, which cuts the polynomial according to the searched speed. In this way, much of the information about the background is suppressed in the calculation of $\tilde{f}^l$, when pedestrians are moving fast. This produces features that are more robust for this kind of movement. Fig. 8 shows an interpretation of the kernel function and its parameter s in relation with the pedestrian's horizontal displacements.

As introduced in Eq. (12), the transformation $\mathrm{T}(t)$ can be induced on the kernel function. From Eq. (18) and considering 2D translation, the transformed kernel function can be written as follows:

$$f_{ij}^l(\check{\mathcal{R}}_k^q)=\sum_{k'=k}^{k+e(s)}\prod_{u_l=0}^{U_l-1}\prod_{v_l=0}^{V_l-1}[\check{r}_q^{k'}(\mathbf{u}_l+\mathbf{t})]^{h(u_l+i,v_l+j)}.$$

Rewriting Eq. (15) for the transformed kernel function yields:

$$\tilde{f}^l(\check{\mathcal{R}}_k^q)=\frac{1}{\check{M}\check{N}}\sum_{i=0}^{\check{M}-1}\sum_{j=0}^{\check{N}-1}f_{ij}^l(\check{\mathcal{R}}_k^q) \tag{19}$$

By defining different kernel parameters $\mathbf{w}_l$, a vector of invariants can be constructed for each ROI $\tilde{\mathbf{f}}(\check{\mathcal{R}}_k^q)=(\tilde{f}^1(\check{\mathcal{R}}_k^q),\ldots,\tilde{f}^L(\check{\mathcal{R}}_k^q))$, with $L\in\mathbb{N}$.

The whole invariant extraction process can be described as a data in-feature out (DAI-FEO) fusion process [36] (see Fig. 9). As the kernel function is given by a polynomial, whose terms collect information about different frames (that is on different times), the calculation of this polynomial can be interpreted as a spatio-temporal fusion given by $\mathbb{R}^{U_l\times V_l\times(e(s)+1)}\to\mathbb{R}$.

4. Detection-classification

An SVM is used for the classification of the invariant vectors $\tilde{\mathbf{f}}_k^q:=\tilde{\mathbf{f}}(\check{\mathcal{R}}_k^q)$. This classifier has a generalization performance equally or significantly better than competing methods [40]. This property is fundamental for the classification of pedestrians, which exhibit a big shape, size and posture variability.

For this work, six classes c have been defined according to pedestrians directions and speeds. The set of all classes is defined as $\mathcal{C}=\{0,1,2,3,4,5\}$, where $c\in\mathcal{C}$. The class corresponding to a certain feature vector $\tilde{\mathbf{f}}_k^q$ is denoted by c_k^q. The class $c=0$ corresponds to a non-pedestrian, i.e., objects that are warm enough to be extracted from the IR video, but that are not persons. The remainder of the classes c from 1 to 5 correspond respectively to a pedestrian who is not moving, walking to the right, walking to the left, running to the right, and running to the left. How to differentiate between walking and running is not strictly defined by gate analysis criteria, but it is based on the manual classification performed by a group of drivers.

Three approaches have been studied for performing the classification of the pedestrians. The simplest method, which will be

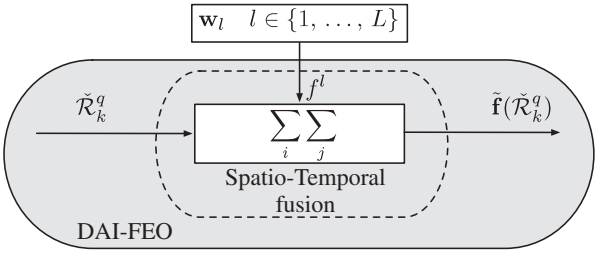


Fig. 9. Invariant extraction.

Please cite this article in press as: A. Pérez Grassi et al., Information fusion to detect and classify pedestrians using invariant features, Informat. Fusion (2010), doi:10.1016/j.inffus.2010.06.002

referred to as approach A, consists in training an SVM with the six described classes (see Fig. 10A). In this case, for each feature vector $\tilde{\mathbf{f}}_k^q$, a classification value $c_k^q \in \mathcal{C}$ is obtained. This process can be described as a feature input-decision output (FEI-DEO) fusion [36]. As will be shown in Section 5, the results of this method are promising.

The second approach, called approach B, looks for an improvement in the detection rate, that is the recognition of non-pedestrians, with respect to approach A. This method includes separating the detection from the classification and in using the output of the first one as an extra input feature for the second one (see Fig. 10B). In this case, a first SVM (detection SVM) is trained only with two classes: pedestrian ($z = 1$) and non-pedestrian ($z = 0$). The output of this SVM is a decision by itself, but for the complete system it is a new feature. Thus, from the complete system point of view, this detection process can be described as feature in-feature out (FEI-FEO) fusion [36]. A feature vector $(z_k^q, \tilde{\mathbf{f}}_k^q)$ forms the input to a second SVM (classification SVM), which was trained with all classes in $\mathcal{C}$. The complete classification method can be described as a FEI-DEO fusion, which is performed by the concatenation of a FEI-FEO and a FEI-DEO fusion process. As will be shown in Section 5, the detection results are improved with respect to approach A, but the classification performance decreases.

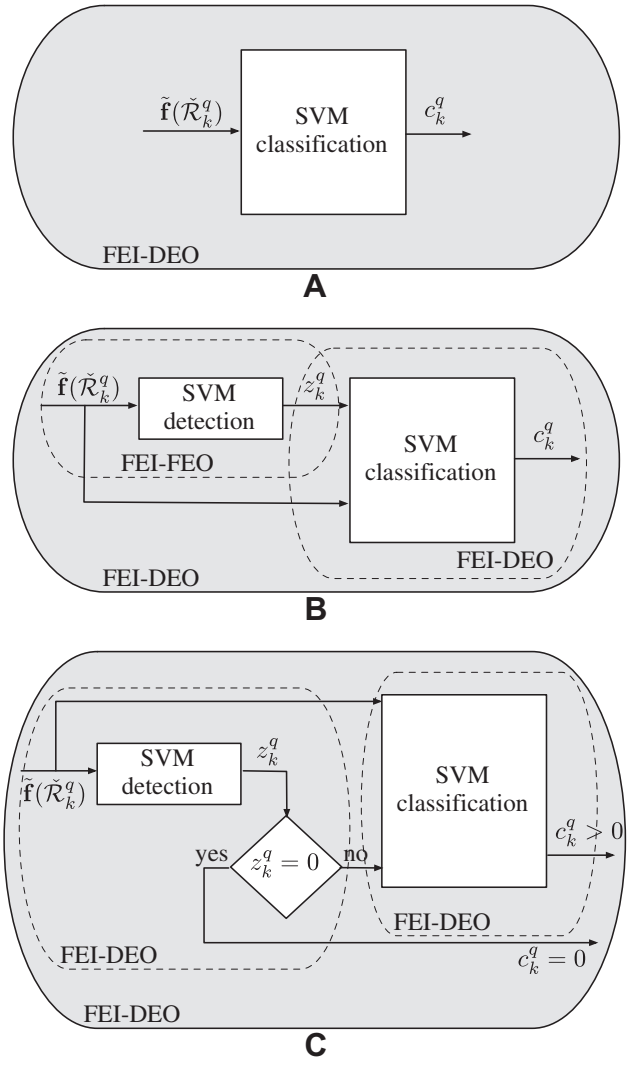


Fig. 10. Feature classification.

The last approach, approach C, integrates the two previous methods to combine both advantages: a high rate of detection and classification. Again, two classifiers are used, a detection SVM and a classification SVM, but this time the detection result z_k^q is considered to be a decision (see Fig. 10C). If the detection is negative $z_k^q = 0$, the feature vector $\tilde{\mathbf{f}}_k^q$ is classified directly as a non-pedestrian $c_q^k = z_k^k = 0$. If, on the contrary, $z_k^q = 1$, then the feature vector $\tilde{\mathbf{f}}_k^q$ is classified by the classification SVM, which is trained with all classes in $\mathcal{C}$. The including of class $c = 0$ in the classification SVM allows to correct false positives of the first classifier. As in the previous approaches, the complete classification system constitutes a FEI-DEO fusion. A comparison of the three proposed classification methods is presented in the following section.

5. Results

In order to train and test the different SVMs, a database of IR videos $\mathcal{G} = \{\mathcal{G}^b | b \in \{1, \ldots, B\}\}$ and lidar signals $\mathcal{D} = \{\mathcal{D}^b | b \in \{1, \ldots, B\}\}$ of real traffic scenes has been recorded. The set of all ROIs extracted from $\mathcal{G}$ and $\mathcal{D}$ is denoted by $\mathcal{R}$. The set of invariants vectors calculated for all ROIs in $\mathcal{R}$ is represented by $\mathcal{F}$.

The presented results are obtained by a $\mathcal{K}$-fold cross validation. The original set $\mathcal{R}$ is partitioned into $\mathcal{K}$ disjoint sets. A single set is retained as validation data, and the remaining $\mathcal{K} - 1$ sets are used as training data. This process is repeated $\mathcal{K}$ times (the "folds"). Finally, the classification results of all folds are averaged. The average classification rate of a class c in a class c' is denoted by $\bar{p}_{c,c'}(\mathcal{F})$. The rate $\bar{p}_{c,c'}(\mathcal{F})$ represents a correct classification if $c = c'$ and a false one if $c \neq c'$. The average percentage of correct classifications considering all classes is represented by $\bar{p}(\mathcal{F}) = \frac{1}{|\mathcal{C}|}\sum_c \bar{p}_{c,c}(\mathcal{F})$. The SVM used in this work is the LIBSVM [41]. Further, the presented results are generated with radial kernel functions, while the SVM parameters are optimized according to the training sets.

For the presented results, $\mathcal{R}$ is constituted by 44 examples of each class. Ten folds are performed using each time 9/10 of $\mathcal{R}$ to train and 1/10 to test. A list of $L = 200$ polynomials has been defined ($\tilde{\mathbf{f}}_k^q \in \mathbb{R}^{200}$). The kernel function parameters are given by $U_l, V_l \in \{3, \ldots, 15\}$, $s = \{-20, 0, -15, 20, 30\}$, $\eta = 20$ and $K - \zeta = 3$. The selection of K determines how much the system will consider the past. The different approaches will be trained and tested with values of K varying from 0 to 5. The K that yields the maximal $\bar{p}(\mathcal{F})$ will be selected.

As described in Section 4, the first proposed classification approach (approach A) includes performing the classification with one SVM (see Fig. 10A). Fig. 11 plots the obtained values of $\bar{p}(\mathcal{F})$ for different values of K. For approach A, the best $\bar{p}(\mathcal{F}) = 86.35\%$ is achieved with $K = 4$. Table 1 shows the values of $\bar{p}_{c,c}$ for all classes using approach A and $K = 4$. Table 2 shows the corresponding con-

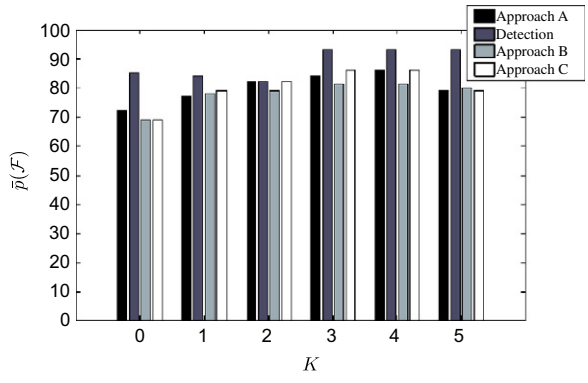


Fig. 11. Average correct classification rate $\bar{p}(\mathcal{F})$ for different values of K and different classification approaches.

Please cite this article in press as: A. Pérez Grassi et al., Information fusion to detect and classify pedestrians using invariant features, Informat. Fusion (2010), doi:10.1016/j.inffus.2010.06.002

fusion matrix. The principal false classification rates are between walking and running classes in the same direction, where the maximum is given by $\bar{p}_{5,3}(\mathcal{F}) = 16.5\%$. Additionally, 7% of non-pedestrians are classified as not moving pedestrians (false positives), and 12% pedestrians are classified as non-pedestrians (false negatives).

Approach B performs the classification in two steps, one for the detection and one for the classification. The detection results are incorporated to the classification as new features (see Fig. 10B). The first step is to train the detection SVM with the classes pedestrian and non-pedestrian. With this objective, a new training list $\mathcal{R}_{det} \subset \mathcal{R}$ is generated with 32 non-pedestrians and 32 pedestrians, where the pedestrians are selected from the different classes $c > 0$ in a proportional way. Now, the best value of K must be selected for the detection SVM. As shown in Fig. 11, $\bar{p}(\mathcal{F})$ gives ≈95% for $K = 3$, 4, 5. Then, the detection is performed with $K = 3$. In this case, the detection SVM gives $\bar{p}_{01}(\mathcal{F}) = 0\%$ false positives and $\bar{p}_{10}(\mathcal{F}) = 5.9\%$ false negatives. The next step is to search the optimal K for the classification SVM. Fig. 11 shows that $K = 4$ yields the best results for the classification SVM. The final values of $\bar{p}_{c,c}(\mathcal{F})$ for this approach are shown in Table 1. There are no misclassifications for the non-pedestrian class, i.e., $\bar{p}_{0,c>0} = 0$. For $c > 0$, the classification performance is lower than for approach A. The resulting classification confusion matrix is shown in Table 3. Although the false negative rate given by the detection SVM was of 5.9% the classification SVM elevates this number considerably. The false positive rate remains zero.

The approach C combines the advantages of the previous methods by using the detection result as a trigger (see Fig. 10C). For the detection SVM all previous results remain valid. As plotted in Fig. 11 for this approach, the best value of K for the classification SVM is 4. The resulting values $\bar{p}_{c,c}(\mathcal{F})$ are shown in Table 1. Approach C achieves a general improvement in the classification performance with respect to approaches A and B. Table 4 shows the confusion matrix. This approach presents a reliable detection rate, with 0% false positives and 5.9% false negatives. The other values of $\bar{p}_{c,c'}(\mathcal{F})$, for $c,c' > 0$, remain in the same order as in approach A. Fig. 12 shows some images and their corresponding classification results. The classification results do not depend on the number of pedestrians. Thus, if no occlusion is considered, the extension of the presented results to more complex scenes is straightforward.

Table 1
$\bar{p}_{c,c}(\mathcal{F})$ for approaches A, B and C.

Class	Approach A (%)	Approach B (%)	Approach C (%)
Non-pedestrian ($c = 0$)	93	100	100
No movement ($c = 1$)	90	89	100[a]
Walking to the right ($c = 2$)	85.5	76	82
Walking to the left ($c = 3$)	78.5	73	77
Running to the right ($c = 4$)	84.5	76	90
Running to the left ($c = 5$)	81	76	83

[a] Detection true positive rate is 94.1%.

Table 2
Classification confusion matrix using one SVM (see Fig. 10A). Values given in percent (%).

	c					
	0	1	2	3	4	5
c'						
0	93	10	0	0	0	2
1	7	90	0	2.5	2.5	2
2	0	0	85.5	0	6.5	2.5
3	0	0	0	78.5	2.5	12.5
4	0	0	14.5	2.5	84.5	0
5	0	0	0	16.5	4	81

Table 3
Classification confusion matrix using an SVM for the detection and one for the classification, where the output of the first one is a feature input for the second one (see Fig. 10B). Values given in percent (%).

	c					
	0	1	2	3	4	5
c'						
0	100	12	13	7	9	5
1	0	89	0	3	3	2
2	0	0	76	3	7	3
3	0	0	0	73	3	13
4	0	0	12	3	76	3
5	0	0	0	13	4	76

Table 4
Classification confusion matrix using an SVM for the detection and one for the classification, where the second SVM is performed only in the case of a positive detection (see Fig. 10C). Values given in percent (%).

	c					
	0	1	2	3	4	5
c'						
0	100	0	0	0	0	0
1	0	100	0	3	3	2
2	0	0	82	0	0	3
3	0	0	0	77	3	13
4	0	0	18	3	90	0
5	0	0	0	18	4	83

5.1. Comparison with HOG

Histograms of Oriented Gradients (HOG) are well-known features used for pattern recognition. They were first applied by Dalal

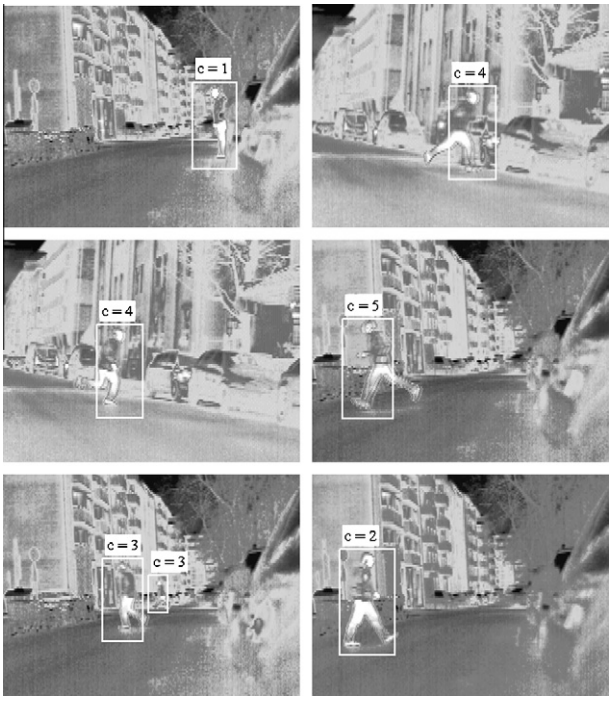


Fig. 12. Classification results.

Please cite this article in press as: A. Pérez Grassi et al., Information fusion to detect and classify pedestrians using invariant features, Informat. Fusion (2010), doi:10.1016/j.inffus.2010.06.002

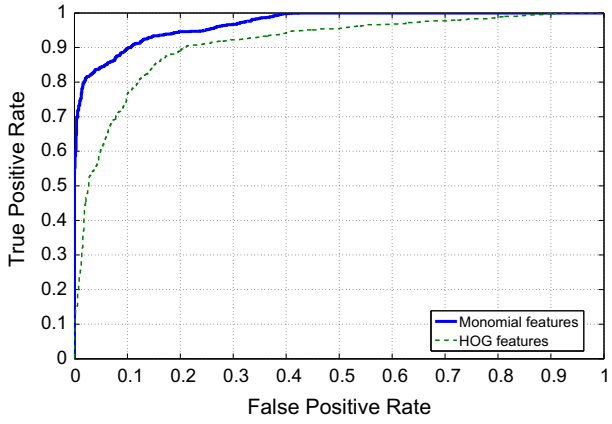


Fig. 13. ROC curve comparing the HOG method with the proposed features.

Table 5
Classification confusion matrix using approach C and HOG features.

	c (%)					
	0	1	2	3	4	5
c'						
0	91	0	3	3	5	0
1	0	88	5	11	5	5
2	2.5	3	61	5	18	5
3	0	5	11	63	3	19
4	6.5	0	12	3	52	10
5	0	4	8	15	17	61

and Triggs in the scope of pedestrian detection [32] and were also used in several other works [22,15,42,43].

Two comparisons are presented. The first one analyzes the detection performance of the proposed features on single gray-level frames. For this purpose, the Daimler Pedestrian Classification Benchmark [44] is used to test both methods. This database consists of grayscale labeled images representing pedestrian and non-pedestrian classes. From this database, 1000 images of each class (pedestrian, non-pedestrian) are used to train the SVM, while a disjoint set of images of the same size is used to test it. For the HOG method, square cells with 20-pixel sides, blocks of two cell and nine histogram bins per cell were used to calculate the features. L^2 norm was selected as a contrast normalization scheme [32]. For the method proposed in this work, 200 features were generated using the same ranges for the kernel function presented in the previous section and $K = 0$. Fig. 13 presents the resulting ROC curve. This curve shows that the proposed method achieves a significantly higher true positive rate with the same rate of false positives as the HOG approach.

The previous analysis involves only the detection performance of the proposed features in gray-level images. Further, the presented method was compared with the HOG method using K frames of IR videos. To adapt the HOG method to the proposed spatio-temporal fusion, the HOG features were calculated on each image separately and then concatenated. The following comparison is based on the approach C described above. The first step consists in finding the best values of K for the detection and classification SVMs when using HOG features. Proceeding as described in Section 5, the best values of K for HOG features are $K = 0$ and $K = 1$ for the detection (91% true negatives and 93% true positives) and classification SVMs respectively. As shown in the previous section, $K = 3$ and $K = 4$ are the best values of K for the proposed invariant features. Table 5 shows the confusion matrix resulting from the HOG features. Comparing these results with those of Table 4, it can be concluded that the presented method improves all classification rates. The most significant improvement is given by the classification rate of pedestrian movements (walking and running in both directions), where the improvement is between 20% and 40%.

6. Conclusion and outlook

This paper presented a new approach to classify pedestrians according to their horizontal movements. The classification is based on the extraction of invariant features — no tracking or movement analysis have been incorporated. The developed method achieves a detection rate of 95%, and classification rates ranging between ≈80% and 95%, which are significantly superior to those achieved with HOG features. These results demonstrate the potential of the proposed method in classifying pedestrians. Additionally, other traffic participants, such as cyclists, children and animals can be incorporated to the classification. Time restrictions as well as the consequences of the sensor displacement remain an open issue, but they can be generally overcome and will be in the scope of future investigations.

Acknowledgments

The authors gratefully acknowledge partial support of this work by the Deutsche Forschungsgemeinschaft (German Research Foundation) within the Transregional Collaborative Research Centre 28 "Cognitive Automobiles."

References

[1] European Road Safety Observatory, Annual statistical report, 2008.
[2] S.J. Krotosky, M.M. Trivedi, On color-, infrared-, and multimodal-stereo approaches to pedestrian detection, in: Transactions on Intelligent Transportation Systems, vol. 8, 2007, pp. 619–629.
[3] M. Bertozzi, A. Broggi, C. Caraffi, M.D. Rose, M. Felisa, G. Vezzoni, Pedestrian detection by means of far-infrared stereo vision, Computer Vision and Image Understanding 106 (2007) 194–204.
[4] Y. Chen, C. Han, Night-time pedestrian detection by visual-infrared video fusion, in: World Congress on Intelligent Control and Automation, 2008.
[5] S. Gidel, P. Checchin, T.C. Christophe Blanc, L. Trassoudaine, Parzen method for fusion of laserscanner data: application to pedestrian detection, in: IEEE Intelligent Vehicles Symposium, 2008.
[6] G. Gate, F. Nashashibi, Using targets appearance to improve pedestrian classification with a laser scanner, in: IEEE Intelligent Vehicles Symposium, 2008, pp. 571–576.
[7] S. Wender, K.C.J. Dietmayer, An adaptable object classification framework, in: IEEE Intelligent Vehicles Symposium, 2006.
[8] M.-M. Meinecke, M.A. Obojski, M. Töns, M. Dehesa, SAVE-U: first experiences with a pre-crash system for enhancing pedestrian safety, in: 5th European Congress and Exhibition on Intelligent Transport Systems, 2005.
[9] B. Fardi, U. Schuenert, G. Wanielik, Shape and motion-based pedestrian detection in infrared images: a multi sensor approach, in: IEEE Intelligent Vehicles Symposium, 2005, pp. 18–23.
[10] L.N. Pangop, S. Comou, F. Chausse, R. Chapuis, S. Bonnet, A bayesian classification of pedestrians in urban areas: the importance of the data preprocessing, in: IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, 2008.
[11] C. Premebida, G. Monteiro, U. Nunes, P. Peixoto, A lidar and vision-based approach for pedestrian and vehicle detection and tracking, in: IEEE Intelligent Transportation Systems Conference, 2007.
[12] Z. Wang, J. Zhang, Detecting pedestrian abnormal behavior based on fuzzy associative memory, in: Fourth International Conference on Natural Computation, 2008.
[13] Z. Chen, D.C.K. Ngai, N.H.C. Yung, Pedestrian behavior prediction based on motion patterns for vehicle-to-pedestrian collision avoidance, in: International IEEE Conference on Intelligent Transportation Systems, 2008.
[14] S. Bota, S. Nedesvchi, Multi-feature walking pedestrians detection for driving assistance systems, IET Intelligent Transport Systems 2 (2008) 92–104.
[15] P. Geismann, G. Schneider, A two-staged approach to vision-based pedestrian recognition using haar and hog features, in: IEEE Intelligent Vehicles Symposium, 2008.
[16] M. Enzweiler, D.M. Gavrila, A mixed generative-discriminative framework for pedestrian classification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2008.

[17] M. Bertozzi, A. Broggi, S. Ghidoni, M.D. Rose, Pedestrian shape extraction by means of active contours, Field and Service Robotics 42 (2008) 265–274.
[18] M. Thuy, A. Pérez Grassi, V.A. Frolov, F. Puente León, Fusion von MIR-Bildern und Lidardaten zur Klassifikation menschlicher Verkehrsteilnehmer, in: M. Maurer, C. Stiller (Eds.), Workshop Fahrerassistenzsysteme, vol. 5, 2008, pp. 168–175.
[19] D.G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2004) 91–110.
[20] B. Leibe, E. Seemann, B. Schiele, Pedestrian detection in crowded scenes, in: CVPR'05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) – IEEE Computer Society, vol. 1, Washington, DC, USA, 2005, pp. 878–885. doi:http://dx.doi.org/10.1109/CVPR.2005.272.
[21] F.H.C. Tivive, A. Bouzerdoum, A biologically inspired visual pedestrian detection system, in: IEEE International Joint Conference On Neural Networks, 2008.
[22] M. Bertozzi, A. Broggi, M.D. Rose, M. Felisa, A. Rakotomamonjy, F. Suard, A pedestrian detector using histograms of oriented gradients and a support vector machine classifier, in: IEEE Intelligent Transportation Systems Conference, 2007.
[23] J. Dong, J. Ge, Y. Luo, Nighttime pedestrian detection with near infrared using cascaded classifiers, in: IEEE International Conference on Image Processing, 2007.
[24] B. Fardi, I. Seifert, G. Wanielik, J. Gayko, Motion-based pedestrian recognition from a moving vehicle, in: IEEE Intelligent Vehicles Symposium, 2006.
[25] Y. Chen, Q. Wu, X. He, Motion based pedestrian recognition, in: Congress on Image and Signal Processing, 2008.
[26] L. Havasi, Z. Szlávik, T. Szirányi, Pedestrian detection using derived third-order symmetry of legs, in: International Conference Computer Vision and Graphics, vol. 32, 2004.
[27] S. Wu, S. Decker, P. Chang, T. Camus, J. Eledath, Collision sensing by stereo vision and radar sensor fusion, in: IEEE Intelligent Vehicles Symposium, 2008.
[28] M. Dimitrijevic, V. Lepetit, P. Fua, Human body pose detection using bayesian spatio-temporal templates, Computer Vision and Image Understanding 104 (2) (2006) 127–139.
[29] P. Viola, M. Jones, D. Snow, Detecting pedestrians using patterns of motion and appearance, in: International Journal of Computer Vision, vol. 63, 2005, pp. 153–161.
[30] E. Seemann, B. Leibe, B. Schiele, Multi-aspect detection of articulated objects, in: CVPR'06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, Washington, DC, USA, 2006, pp. 1582–1588. doi:http://dx.doi.org/10.1109/CVPR.2006.193.
[31] S. Milch, M. Behrens, Pedestrian detection with radar and computer vision, 2001.
[32] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: C. Schmid, S. Soatto, C. Tomasi (Eds.), International Conference on Computer Vision & Pattern Recognition, vol. 2, 2005, pp. 886–893.
[33] A. Mohan, T. Poggio, Example-based object detection in images by components, in: Proceedings of IEEE Transactions on PAMI, vol. 23, 2001, pp. 349–361.
[34] F. Xu, X. Lui, K. Fukimura, Pedestrian detection and tracking with night vision, in: Proceedings of IEEE Transactions on Intelligent Transportation Systems, vol. 6, 2005, pp. 63–71.
[35] M. Thuy, F. Puente León, Non-linear, shape independent object tracking based on 2d lidar data, in: Intelligent Vehicles Symposium, 2009 IEEE, 2009, pp. 532–537. doi:10.1109/IVS.2009.5164334.
[36] B. Dasarathy, Sensor fusion potential exploitation-innovative architectures and illustrative applications, Proceedings of IEEE 85 (1997) 24–38.
[37] H. Schulz-Mirbach, Anwendung von Invarianzprinzipien zur Merkmalgewinnung in der Mustererkennung, Ph.D. thesis, Technische Universität Hamburg-Harburg, 1995.
[38] R. Lenz, Group Theoretical Methods in Image Processing, Lecture Notes in Computer Science, Springer, 1990.
[39] J. Wood, Invariant pattern recognition: a review, Pattern Recognition 29 (1) (1996) 1–17.
[40] C. Burges, A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery 2 (1998) 121–167.
[41] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines, Software available at http://www.csie.ntu.edu.tw/cjlin/libsvm, National Taiwan University, 2001.
[42] Q. Zhu, S. Avidan, M.C. Yeh, K.T. Cheng, Fast human detection using a cascade of histograms of oriented gradients, in: CVPR, IEEE Computer Society, 2006, pp. 1491–1498.
[43] H.-X. Jia, Y.-J. Zhang, Fast human detection by boosting histograms of oriented gradients, in: Fourth International Conference on Image and Graphics, ICIG 2007, 2007, pp. 683–688.
[44] S. Munder, D. Gavrilla, An experimental study on pedestrian classification, in: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, 2006, pp. 1863–1868.

Please cite this article in press as: A. Pérez Grassi et al., Information fusion to detect and classify pedestrians using invariant features, Informat. Fusion (2010), doi:10.1016/j.inffus.2010.06.002