Neurocomputing 300 (2018) 17–33

Contents lists available at ScienceDirect

Neurocomputing

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

Computer vision and deep learning techniques for pedestrian detection and tracking: A survey

Antonio Brunetti[a], Domenico Buongiorno[a], Gianpaolo Francesco Trotta[b], Vitoantonio Bevilacqua[a,*]

[a] Department of Electrical and Information Engineering (DEI), Polytechnic University of Bari, Italy
[b] Department of Mechanics, Mathematics and Management (DMMM), Polytechnic University of Bari, Italy

ARTICLE INFO

Article history:
Received 15 September 2017
Revised 10 December 2017
Accepted 8 January 2018
Available online 10 March 2018

Keywords:
Pedestrian detection
Human tracking
Deep learning
Convolutional neural network
Machine learning
Artificial neural network
Features extraction

ABSTRACT

Pedestrian detection and tracking have become an important field in the computer vision research area. This growing interest, started in the last decades, might be explained by the multitude of potential applications that could use the results of this research field, e.g. robotics, entertainment, surveillance, care for the elderly and disabled, and content-based indexing. In this survey paper, vision-based pedestrian detection systems are analysed based on their field of application, acquisition technology, computer vision techniques and classification strategies. Three main application fields have been individuated and discussed: video surveillance, human-machine interaction and analysis. Due to the large variety of acquisition technologies, this paper discusses both the differences between 2D and 3D vision systems, and indoor and outdoor systems. The authors reserved a dedicated section for the analysis of the Deep Learning methodologies, including the Convolutional Neural Networks in pedestrian detection and tracking, considering their recent exploding adoption for such a kind systems. Finally, focusing on the classification point of view, different Machine Learning techniques have been analysed, basing the discussion on the classification performances on different benchmark datasets. The reported results highlight the importance of testing pedestrian detection systems on different datasets to evaluate the robustness of the computed groups of features used as input to classifiers.

1. Introduction

The growing interest in autonomous cars demonstrated by the huge investments made by the biggest automotive and IT companies [1], as well as the development of machines and applications able to interact with persons [2–13], is playing an important role in the improvement of the techniques for vision-based pedestrian tracking. In fact, autonomous machines able to act in not-controlled environments represent an high risk for any person who may be in their range of action.

In 2015, in the United States, more than 5000 pedestrians were killed due to traffic crashes [14]: one pedestrian dies every 1.6 hours due to car accident. Additionally, in the same year, almost 130000 pedestrians were treated in emergency departments for non-fatal crash-related injuries. Pedestrians are 1.5 times more likely than passenger vehicle occupants to be killed in a car crash on each trip [14–17]. The statistics reported in [14] state alarming numbers for EU too, even though the general trend of pedestrians' deaths is reducing thanks to the introduction of driving supports, such as auto-breaking system. For these reasons, in the last decades, people detection and tracking has become an important research area in computer vision.

From 1990 to 2016, scientific community has shown an ever-growing interest in human detection and tracking. As reported in Fig. 1, more than 5000 publications in this topic have been published and indexed in *Web of Science*, ranging from human detection to pedestrian tracking using 2D and 3D vision systems, or considering indoor and outdoor environments.

Some other surveys regarding pedestrian detection have been presented in the literature so far. In [18] and [19] the authors focused the topic of the survey on a taxonomy of system functionalities considering the structure of the motion capture system and the different information to be processed.

Solichin et al. have focused the work on the steps needed in the process of pedestrian detection, including input devices, datasets and methods for detection and, finally, on some open issues related to pedestrian detection [20].

Zhou and Hu have written a survey on the human detection and tracking systems from a clinical and diagnostic point of view,

* Corresponding author.
E-mail address: vitoantonio.bevilacqua@poliba.it (V. Bevilacqua).

https://doi.org/10.1016/j.neucom.2018.01.092
0925-2312/

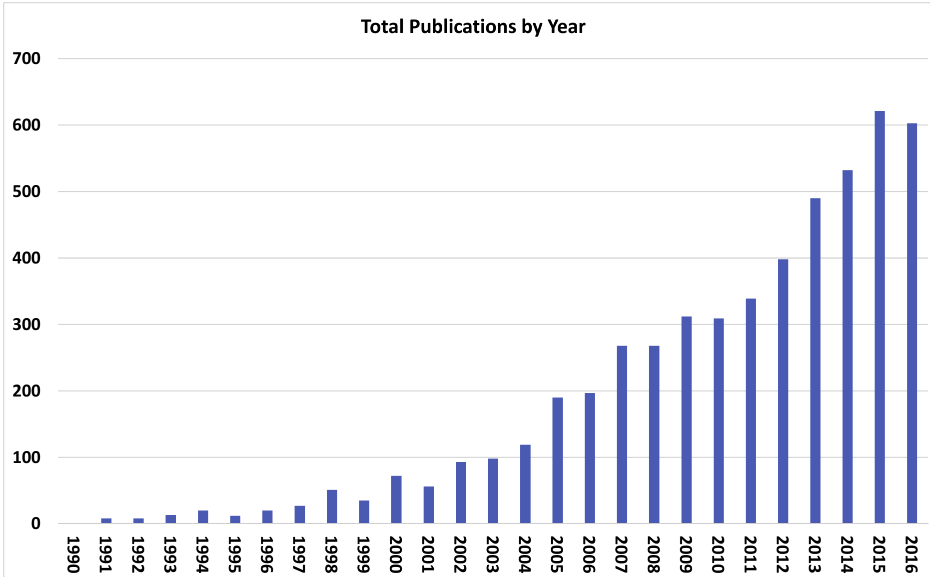

Fig. 1. Total Publications from 1990 to 2016 with Keyword: Human Detection and Tracking - Source:*Web of Science*.

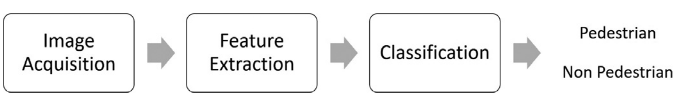

Fig. 2. Steps needed for pedestrian classification following a features-based model.

Fig. 3. Steps needed for pedestrian classification following a model based on Deep Learning strategy.

highlighting the differences between visual tracking (marker-based or marker-less) and non-visual tracking using magnetic sensors, inertial sensors and others [21].

In [22], the authors have presented a survey concerning monocular pedestrian detection systems focusing on the methodologies for the selection of Regions Of Interest (ROIs), classification methods and tracking.

In [23] and [24], the authors have discussed two surveys focused on pedestrian detection and tracking systems related to the Pedestrian Protection Systems (PPSs). Specifically, while in the first survey the authors consider and review general pedestrian detectors, in the latter the authors focus only on systems dedicated to PPSs.

Dollar and his colleagues [25] have focused on the main methods for pedestrian detection in monocular images performing an accurate ranking on benchmark datasets, while in [26] the authors have collected and reviewed some works, marginally introducing deep architectures.

The above-mentioned surveys report the state-of-the-art about pedestrian detection and tracking systems in terms of acquisition technologies, e.g. 2D and 3D configurations, and processing methodologies; however, recent adoption of Deep Learning (DL) methodologies and, in particular, Convolutional Neural Networks (CNNs) for pedestrian detection and tracking deserves a dedicated state-of-the-art survey.

Generally, the process of vision-based pedestrian detection can be considered constituted by three fundamental steps, as depicted in Fig. 2: (i) Image Acquisition, (ii) Feature Extraction and (iii) Classification. As will be discussed in the following sections, the introduction of DL architectures, or deep structures inspired to the human visual cortex, in the context of object recognition, allowed the removal of the feature extraction step (Fig. 3), preserving the other ones.

As pointed out by the reported figures, the two approaches differ for the removed step only. However, the extraction of features is not completely removed from the workflow, but it is an automatic procedure performed by the deep classifier which is generally constituted by several processing layers that, taking images as input, compute features at different layers of abstraction [27–30]. In this way, the design of a such a kind classifiers is considerably simplified since the design of procedures for the extraction of the so called "hand-crafted features", able to perform an accurate classification, is the most difficult step. Nevertheless, the incorporation of feature extraction in the classification process, allowing a faster run-time execution, lead to a longer training time respect to the hand-crafted features based approach [31].

Video tracking is a complex process which allows to locate and follow single or multiple objects over time using several sensors. Due to the need of a remarkable improvement in both acquisition and processing systems, a lot of works dealing with tracking could be found in literature. In fact, each moving object in the world could be potentially tracked regardless the tracking system. For example, complex systems based on radar or GPS are widely studied and used currently in different contexts, e.g. aviation industry or ground movements tracking [32–36].

Among the variety of traceable objects, human tracking is the most interesting since the processes of human detection and segmentation in images and videos are difficult due to the large variety of conditions and variables to take into account for this task, besides the well-known problems related to images segmentation, such as noise [37–41].

The automatic tracking of humans in video has always been an interesting research topic, as it is a cross-domain research area with infinite applications in different fields. In fact, the potential applications of human motion capture led to the development of systems in several domains, such as surveillance, control, and analysis. In addition, there are some recent research fields where the automatic tracking of humans in video sequences is rising up, such as human-computer interaction and augmented reality [8,36,42–46].

Regardless of the kind of object to be tracked, the identification of Regions Of Interest (ROIs) is the first and most important step in the most of computer vision applications, including object tracking. This step requires the application of some image processing techniques in order to make easier the identification and selection of ROIs; the difficulty of this approach mostly depends on both the acquisition system (e.g., camera resolution, field of view and technology) and the environmental conditions (e.g., lighting conditions). Although it seems a trivial process, in some approaches the previous sequence of steps could be sufficient to track one or more objects into a video sequence under certain conditions [47–49].

In more complex applications, some features need to be extracted in order to describe the identified regions. The extracted features, whose kind is related to the acquired signal, are then used as input in the subsequent step for the discrimination of the identified objects; finally a tracker is necessary to follow the considered object (or class of objects) during the video flow [3,8,12,13,18,19,21–23,25,26,32,36,43,44,50–99].

To simplify and strengthen the step of ROIs identification, and consequently the overall tracking system, some authors introduced active and passive markers (or optical references) to be applied to the object to track. In literature, several kinds of marker could be found; their nature is strictly related to the technology of the acquisition system, allowing an accurate tracking in several conditions [9,90,100–109].

In recent years, the spread of innovative techniques based on Deep Learning has prompted many research groups to apply these techniques for the segmentation of objects in images and tracking in videos with different aims [110–119]. This kind of approach seems to be very interesting and powerful since the steps needed for the features extraction from the segmented ROIs is overcame thanks to DL architectures that make use of deep classifiers, such as Convolutional Neural Networks (CNNs).

In the sections that follow, we present the application fields of pedestrian detection and tracking systems, first. We then describe the different configurations of the vision systems in the Section 3. Subsequently, we present the different methods for video processing and features extraction in the Section 4 focusing on pedestrian subjects in the Section 4.1. Then, we introduce the approaches pedestrian classification using Machine Learning and Deep Learning techniques in the Section 5, while we present a final discussion in the Section 6. Finally, we present conclusions.

2. Applications

The growing interest for vision-based tracking systems can be explained by multiple factors. To the authors' opinion, the most important factor is the advancement of the related fields that make use of the tracking techniques. In addition, recent researches with background in Artificial Intelligence (AI), Augmented Reality (AR) and medical imaging, as well as the diffusion of low cost video acquisition systems and more powerful processing devices, have contributed to the diffusion of researches in tracking systems.

The three major application areas individuated by the authors are: surveillance, human-machine interaction, and analysis.

2.1. Video surveillance

Surveillance applications based on human tracking are the most diffused in literature. The main goal is the detection of one or more people in the scene for tracking their movements in video flow over time. For example, several systems are able to monitor parking lots, airports or crowded places (Fig. 4).

The main differences that characterize the works found in literature about video surveillance applications consist in the acquisition systems (e.g., colour-space and resolution), the number of potentially traceable subjects (e.g., mono or multi target), and object categorization [51,61,65,77–79,91]. Among the video surveillance applications, the most afforded research topic is focused on pedestrian tracking (see Section 4.1 for more details) [94,97–99,116].

2.2. Human-machine interaction

The human-machine interaction area relates to tasks where the captured human motion is used to provide controlling functionalities for remote controlling and designing virtual game interfaces, virtual environments and animations.

Moreover, tracking systems have been also applied in the entertainment industry where the control of personalized graphic models is making the productions/products more realistic.

In recent years, the interest in using Unmanned Aerial Vehicles (UAVs) to accomplish a series of tasks that can be uncomfortable or dangerous to be performed by human beings has been subject to a constantly increase (Fig. 5). This can be mostly explained by the higher possibility to purchase a cheaper drone, especially for game and sport purposes. This has pushed the scientific community to investigate the capabilities of UAVs in video-based tracking applications [7,9,12,13,90,91,94,96–99,101–103,105,120–126].

Besides the UAV control, a lot of research and investments have been done by big corporations to support the research in self-driving cars [1,3,92].

2.3. Analysis

The analysis of captured motion data may be used in different clinical studies, e.g. to diagnose orthopaedic diseases, to help athletes in understanding and improving their performance, to restore patients' functional capability in stroke rehabilitation or to prevent fall accidents [21,44,95,100,106–109]. In this kind of applications, the patients' activities need to be continuously monitored, and subsequently corrected during motor-rehabilitation sessions [127–129].

Furthermore, these types of applications are used to answer questions about what people are doing and where and when they act. To achieve this goal, the algorithms implemented in these applications build people's appearance patterns and trace people with relative identity (who) through occlusion events in the imagery. So they are used to increase awareness of security issue by performing analysis of actions, activities and behaviors both for crowds and individuals; for example, such systems are used for queue and shopping behavior analysis, detection of abnormal activities, and person identification [18,19,51–53,55,62,69,74,90,104,130].

3. Vision systems for pedestrian detection

The systems used to capture human motion consist of subsystems for sensing and processing, respectively. The operational complexity of these subsystems is typically related, i.e. the more complex the previous step is, the simpler the following one will be and vice versa. This trade-off between the complexities also relates to the use of active versus passive sensing.

Active sensing operates by placing devices on the subject and in the surroundings which transmit or receive generated signals. Active sensing allows for simpler processing and is widely used when the applications perform in well-controlled environments; for example, the most of applications in the analysis and control areas make use of active sensing (Fig. 6).

Passive sensing is based on "natural" signal sources, e.g. visual light or other electromagnetic wavelengths, and generally requires

Fig. 4. A representation of pedestrian detection system in outdoor environment. Boxes rounding pedestrian show the correct detection of person in different poses. Contribution from [25].

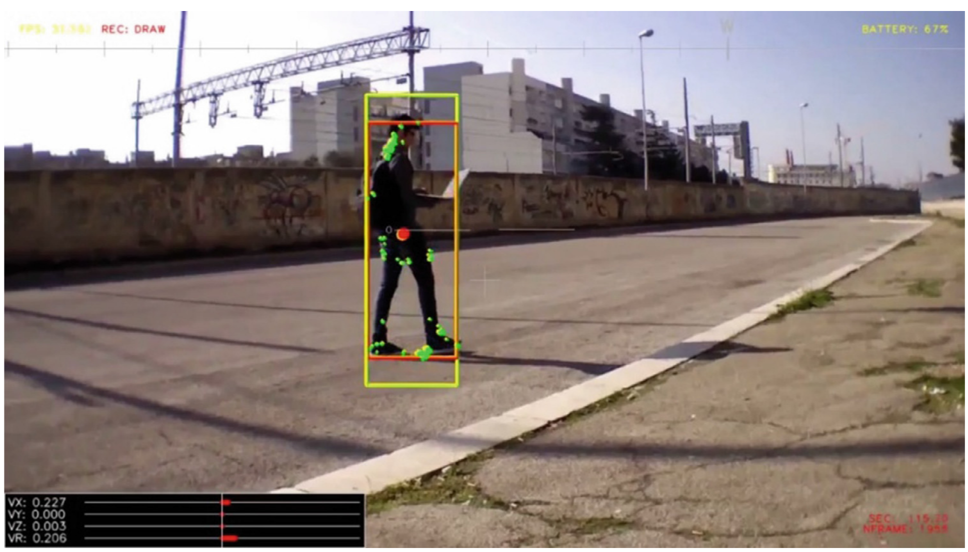

Fig. 5. An application of drone following human Key points are shown on pedestrian; in the bottom-left are shown the inputs to control the drone's trajectory. Contribution from [12].

no wearable devices. An exception can be make if markers are attached to the subject for an easier motion capture process. Visual markers are not as intrusive as the devices used in active sensing where passive sensing is mainly used in surveillance and some control applications where mounting any kind of active device on the subject it is not allowed.

Computer vision applications based on the passive sensing approach have challenged active sensing within all the considered application areas. Even though the use of markers could be a good compromise between passive and active sensing, the application of any kind of passive or active marker in real situations for tracking objects in uncontrolled or random environments is generally inconvenient or often impossible. For these reasons, systems able to detect and track objects considering only the acquired images from the vision system are needed.

Considering passive sensing, currently the great majority of the algorithms that accomplish similar tasks relies on colour information or on the use of external devices that track the target position [121,124–126].

Regardless of the use of active or passive sensors, vision systems for pedestrian detection may be differentiated considering the video acquisition technology (2D vs 3D), or the environmental conditions (Indoor vs Outdoor). These two topics will be afforded in the following two sections.

3.1. 2D vs 3D

Video acquisition technology is one of the fundamental aspects that concern with pedestrian, and generally, object detection and tracking. In literature, most of the works dealing with

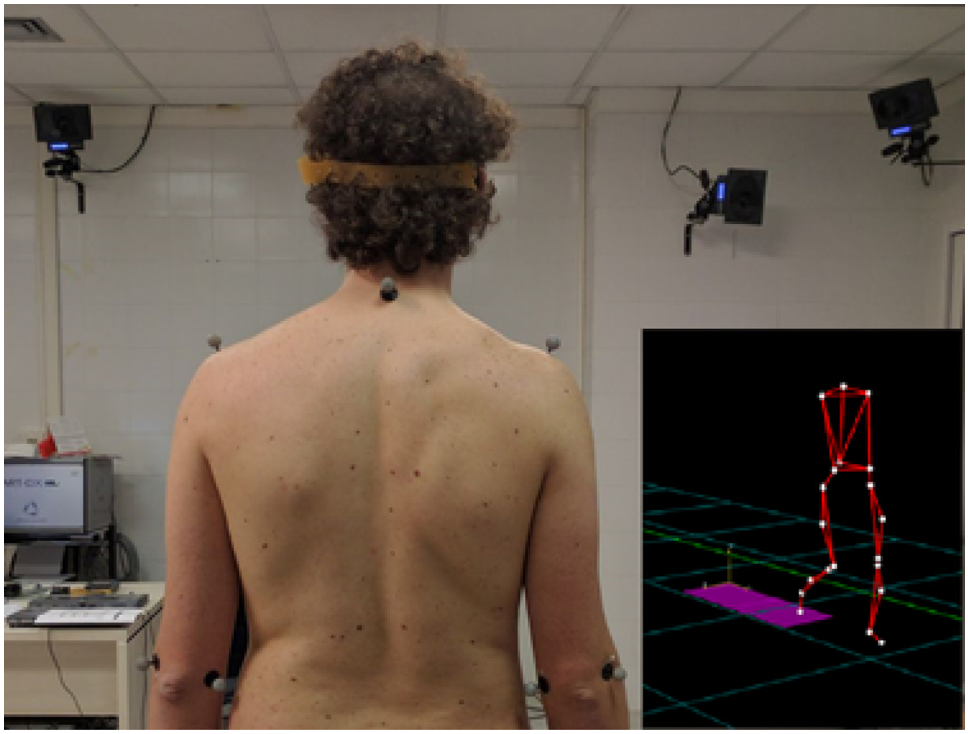

Fig. 6. A 3D acquisition system for gait analysis. The 3D skeleton is reconstructed thanks to the application of visual markers on the person. Contribution from [107].

pedestrian detection use a 2D acquisition system and Machine Learning (ML) techniques to perform a large variety of tasks [12,13,21–23,25,26,50,51,54,56,59–64,67–69,71,72,74–81,84–91,93,94,96–99,131–134].

In the most of cases, a 2D video is sufficient for pedestrian detection since videos contain extremely valuable information that can be extracted after an appropriate processing, i.e. the 2D coordinates of a detected person. Moreover, as will be discussed in the following section (Section 4), motion information, which is generally extracted using a 3D vision system, could be obtained following a bi-dimensional approach [3,18,19,21,44,52,53,55,65,76,92,95,135,136].

Conversely, in applications where the motion of a person should be acquired with high levels of accuracy, like pedestrian protection systems for autonomous vehicle or clinical and diagnostic environments for gait analysis, 3D motion capture systems are needed since 2D acquisition could lead to an excessive loss of information. 3D systems are able to generate a pedestrian coordinates representation in the 3D space (*x, y, z* planes of motion), often used to generate a representational model in virtual environments. This kind of systems generally differ for the sensing device that affect the overall cost of the application; in particular, 3D systems may use RGB-D cameras, which consist in the combination of a classic RGB camera and a Depth camera based on infra-red (IR) light acquisition [137,138]; stereo camera which is a type of camera constituted by two or more lenses with a separate image sensor or film frame for each lens to simulate human binocular vision, and therefore gives it the ability to capture three dimensional images, by using stereo photograph. Multi-camera system allows to capture the scene from several points of view and needs, in the latter case, calibration and registration phases and optical markers to obtain high levels of accuracy during people tracking.

Typically, 3D optical systems are more expensive than 2D ones, but 2D video cameras are much easier to use and faster to configure and set-up. On the contrary, using a 3D acquisition system, the detection and tracking of people could be easier and more accurate than using 2D acquisition system.

3.2. Indoor vs outdoor

The environmental conditions are the second aspect to take into account during the design of a new system for pedestrian detection and tracking. In fact, illuminating conditions, as well as the variability of subjects in the scene are critical aspects to be considered. Since outdoor environments generally have tricky boundary conditions which need performing algorithms for their handling, the most of literature on pedestrian detection and tracking systems concerns with applications for outdoor. On the other hand, since the conditions in indoor environments are more controllable, there are applications which show robust and performing human tracking.

Excluding extremely variable indoor crowded places, such as airports, that are considered the same as outdoor, the most of works on pedestrian detection in indoor areas use external markers to detect a moving object inside the scene [21,90,100,104,106–109].

Visual markers are useful easy-to-locate tools and thus allow an easier tracking within a video stream in a controlled environment. Thanks to their non-invasive nature, they could be easily applied on the object to be tracked into a scene. Both the detection and tracking systems are simplified since common RGB cameras could be used to acquire the scene; in particular, high performance pattern matching systems may be used to find the marker inside each frame.

The use of markers in indoor environment may have multiple aims; for example, while in [104], Mehner et al. placed a marker on each person's head that could be recorded in the scene and have used it to track the different trajectories for subsequent analyses, Naseer et al. describe a system to follow pedestrians using a quadcopter and use markers to support the motion of the quadcopter itself [90]. In particular, the authors set up two cameras, first one for determining the 3D position of the UAV based on markers placed on the ceiling of a controlled room and second, a depth camera, for detecting a person in the 3D space. The image resulting from the depth camera is then warped, based on the calculated 3D position.

Considering scenarios of human tracking, marker-based systems are highly recommended in applications where the body position needs to be quickly and accurately tracked, while the human skeleton makes unpredictable and complicated motion trajectory. In addition, cluttered scenes, or varied lighting, most likely distract visual attention from the real position of a marker. Given these problems, visual marker-based tracking is preferable.

In these circumstances, simple human tracking is not sufficient; in fact, some applications require to detect and track single body parts, especially in applications within the analysis domain [21,100]. In order to reach high levels of accuracy in human body tracking, it could be necessary to change the marker technology and, consequently, the acquisition system. For example, in [106–109] the authors have used multiple IR cameras with specific visual markers to track human body parts with high accuracy for specific tasks.

From a technical point of view, marker-based tracking systems are easy to implement since, as already stated, markers are employed in controlled environments in terms of lightning and field of view. Scientific community, instead, spent much time to study and implement human detection and tracking system that do not use any kind of marker. In fact, regarding indoor detection, literature reports some recent works describing application were human tracking is performed in indoor environments without any adoption of markers [135,136].

Generally, outdoor pedestrian detection is a more difficult task than indoor one, since the external environment is generally influenced by so many variables and the scenes to be acquired are completely unpredictable. Recent literature contains several works dealing with marker-based human tracking, the most of which make use of drones [51,65,77–79,91,94,96,97,101–103,105,120–123,126,139,140]. In some cases, they do not use external markers to detect and track a human. In [101], the authors have developed a 3D object following system based on visual information acquired from the UAV camera; in particular, the authors recognize a specific object placed into the scene and use this information to control the movement of a drone. In [105], Vasconcelios et al. have used shirts with markers to track a person with a drone; in particular, the authors have developed a "behavioral marker" composed of two different parts: the first one, which is constant, is used for detection and tracking processes; the second part, instead, is variable and is used to adapt the drone behavior to the specific recognized person so that the UAV is able to know which person is targeting and following.

4. Computer vision methods for pedestrian detection

The initial approach for detection and tracking of moving objects into a video flow acquired by a static camera consisted in the Background Subtraction (BS); this technique allows the detection and distinction of moving objects inside a scene using an appropriate background model [48]. Even though algorithms based on BS are quite simple to implement, this approach is not robust to illumination variability, dynamic background, shadows or noise limiting its usage mostly in controlled environments [139–141].

In recent years, a huge number of algorithms have been developed and tested to perform human detection and tracking, but the most of them are based on the following approaches for features extraction and detection:

- *Histograms of oriented gradients* [56]: this method is based on the idea that local object (human or not) appearance and shape can often be characterized considering local intensity gradients or edge directions distribution (Fig. 7). Each video frame is divided into small regions and a local 1-D histogram of gradient directions or edge orientations over the pixels of the block is computed. An improved version of this algorithm, which is able to handle with problems related to illumination or shadowing, is used for the normalization of the histograms considering a group of smaller blocks. In both cases, each generated histogram is considered as image representation and a cascade of classifiers is used to discriminate each sub-region.
- *Haar-like features* [50]: with this approach, the wavelet representation is used to capture the structural similarities between various instances of the class of humans [142]. In particular, 2-dimensional Haar wavelets include basis functions which capture change in intensity along the horizontal, vertical and diagonals (or corners) directions (Fig. 8). As in previous case, each representation is used as input to a classifier. Improved versions of the algorithm are applied to support multi-scale detection.
- *Viola–Jones features* [54]: this approach is an extended version of the rectangle filters presented by Viola and Jones for the static face detection [143,144]; in particular, this approach considers particular filters based on Haar wavelets. In this case, the proposed approach take into account both motion and intensity information even considering sequences of frames.
- *Texture* [145]: features extraction from texture is a quite simple approach and consists in the elaboration of its distribution in the image; in literature several works dealing with textural features extraction could be found [59,145,146]. On the contrary, the classification of pedestrian considering textural features only is a challenging problem due to the high variability of classes to be considered, e.g. pedestrians variations due to clothing and varying lighting conditions. In order to avoid this, textural features are generally used in combination with other kinds of features, such as shape, colour and others.
- *Local Binary Pattern (LBP)* [68]: this technique allows to describe images based on their texture by opportunely considering the neighborhood of each pixel [147]. LBP approach have become very popular due to its robustness against variations in pose or illumination than other methods. As reported in [68], LBP feature vectors are very often used in combination with HOG features to reach higher performance in pedestrian detection.

In the following paragraph, more details about pedestrian detection will be given related to some innovative works making use of these algorithms or their variations.

4.1. Pedestrian detection and tracking

In this paragraph, the most important works dealing with the task of pedestrian detection and tracking are investigated. In Table 1, the performance of each detector are reported in terms of Log-Average Miss Rate (MR) on the most common benchmark databases, namely Inria [56] and Caltech [25,148], along with details about both the detector and the classifier family used in each work.

In [54], Viola et al. describe a pedestrian detection system that integrates image intensity information with motion details; in particular, the authors combined Haar-like features with motion information that were computed considering two consecutive frames in a video sequence. The authors applied the face detector described in [143] and [144] to the pedestrian detection problem, but their results on the benchmark databases show an high Log-Average Miss Rate. Classification was performed considering a sequence of AdaBoost classifiers.

(a) (b)

Fig. 7. HOG features extraction for pedestrian detection. The input image is on the left (a); the output image on the right shows the superimposition of HOG descriptors on the input image (b).

Table 1
Log-Average Miss Rate for some works dealing with pedestrian detection. The implemented detector, the dataset used for training and test, and the classifier are reported.

Detector	*Training set*	*Classifier*	*Test set*	*Log-Average Miss Rate*(%)
Informed Haar [93]	Caltech	AdaBoost	Caltech	34.60
Informed Haar [93]	Inria	AdaBoost	Inria	14.43
VJ [54]	Inria	AdaBoost	Caltech	94.73
VJ [54]	Inria	AdaBoost	Inria	72.48
HOG [56]	Inria	linear SVM	Caltech	68.46
HOG [56]	Inria	linear SVM	Inria	45.98
Shapelet [60]	Inria	AdaBoost	Caltech	91.37
Shapelet [60]	Inria	AdaBoost	Inria	81.70
MultiFtr+CSS [71]	Inria	AdaBoost	Caltech	60.89
MultiFtr+CSS [71]	Inria	AdaBoost	Inria	24.74
MultiFtr+Motion [71]	TUD-Motion	linear SVM	Caltech	50.88
HikSvm [63]	Inria	HIK SVM	Caltech	73.39
HikSvm [63]	Inria	HIK SVM	Inria	42.82
HogLbp [68]	Inria	linear SVM	Caltech	67.77
HogLbp [68]	Inria	linear SVM	Inria	39.10
LatSvm-V1 [64]	Pascal	latent SVM	Caltech	79.78
LatSvm-V1 [64]	Pascal	latent SVM	Inria	43.83
LatSvm-V2 [72]	Inria	latent SVM	Caltech	63.26
LatSvm-V2 [72]	Inria	latent SVM	Inria	19.96
ChnFtrs [67]	Inria	AdaBoost	Caltech	56.34
ChnFtrs [67]	Inria	AdaBoost	Inria	22.18
FeatSynth [75]	Inria	linear SVM	Caltech	60.16
FeatSynth [75]	Inria	linear SVM	Inria	30.88
MultiResC [73]	Caltech	latent SVM	Caltech	48.45
CrossTalk [80]	Inria	AdaBoost	Caltech	53.88
CrossTalk [80]	Inria	AdaBoost	Inria	18.98
VeryFast [81]	Inria	AdaBoost	Inria	15.96
SketchTokens [84]	Inria	AdaBoost	Inria	13.32
Roerei [85]	Inria	AdaBoost	Caltech	48.35
Roerei [85]	Inria	AdaBoost	Inria	13.53
AFS+Geo [86]	Inria	linear SVM	Caltech	66.76
DBN-Isol [82]	Inria	DeepNet	Caltech	53.14
DBN-Mut [88]	Inria	DeepNet	Caltech	48.22
ACF+SDt [89]	Caltech	AdaBoost	Caltech	37.34

In [56], Dalal and Triggs studied the question of feature sets for robust visual object recognition, introducing the Histogram of Oriented Gradient (HOG) features. After reviewing existing edge and gradient based descriptors, the authors showed experimentally that grids of HOG descriptors significantly outperform existing feature sets for human detection. A linear SVM was adopted for human detection and classification; Gaussian SVM was explored too, but run-time result does not perform better than using linear SVM.

In [60], the authors addressed the problem of detecting pedestrians in static images introducing a set of features called "Shapelet". These are a combination of low-level features, which consisted primarily in the gradient responses in images, and then in a set of features automatically learned using an AdaBoost classifier. Finally, another AdaBoost classifier was trained to discriminate between pedestrian and non-pedestrian using Shapelet features as input. The reported results show that the developed approach performs better on Caltech database than on Inria one.

In [63], Maji et al. discussed that it is possible to build histogram intersection kernel SVMs (IKSVMs) with a logarithmic run time complexity considering the number of support vectors as opposed to linear used as standard approach. The authors introduced a variant of HOG features based on a multi-level version of HOG descriptors. They showed that by pre-computing auxiliary tables, it was possible to design an approximate classifier with constant

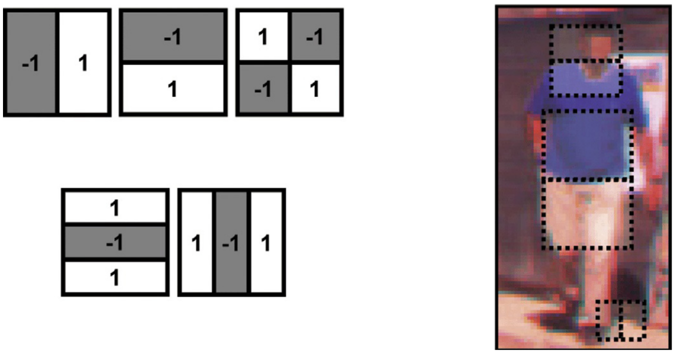

Fig. 8. A representation of some filters from the extraction of Haar-like features from images. A contribution from [76].

runtime and space requirements, independent of the number of support vectors, with negligible loss in classification accuracy on various tasks.

In [64], Felzenszwalb et al. described an approach based on part model for object detection. The authors evaluated HOG features at different levels of resolution leading to "HOG features pyramid", thus allowing the detection of parts that could be moved respect to the detection window. The authors combined a margin-sensitive approach for data mining hard negative examples with a formalism called latent SVM which leads to a non-convex training problem. However, a latent SVM is semi-convex and the training problem becomes convex, once latent information was specified for the positive examples. In [72], Felzenszwalb et al. reduce the dimensionality of the dataset used in [64] through the PCA algorithm. An improved version of the multi-scale detection, together with PCA for dimensionality reduction, led to an improvement of the performance on both the benchmark test sets Inria and Caltech.

In [68], Wang et al. proposed a novel human detection approach capable of handling partial occlusion. In details, a new feature set was introduced considering HOG and LBP features. In order to handle partial occlusions, two detectors were combined: the first is performed globally on the image, while the second (part detector) is executed in ambiguous areas to refine the detection. For each ambiguous scanning window, an occlusion likelihood map was constructed by using the response of each block of the HOG feature to the global detector. The occlusion likelihood map was then segmented by Meanshift approach [149]. The segmented portion of the window with a majority of negative response is inferred as an occluded region. Thanks to this this approach, based on the augmented HOG-LBP feature and the global part occlusion handling method, they achieved very high levels of detection rates considering linear SVM classifiers.

In [67], Dollar et al. studied the performance of "integral channel features" for image classification task, focusing in particular on pedestrian detection. In details, multiple representation of the same input image could be computed applying linear and non-linear transformations. Considering the features extracted from each representation, such as local sums, histograms, and Haar features and their different generalizations, the integral image is then computed [150] and used as input in the classification step. Performance was tested considering three different classifiers: AdaBoost, RealBoost and LogitBoost. In [80], Dollar et al. also investigated the correlations between detector responses at nearby location and scales in an application where cascades help to make sliding windows object detection fast, nevertheless, computational demands remain prohibitive. In particular, the authors selected a restricted subset of features from the group reported in [67], focusing their work on a low-level optimization, leading to an improvement at both compile and run time as well.

In [71], Walk et al. showed that motion features derived from optical flow, if implemented correctly, yield substantial improvements on image sequences, even in presence of low-quality video sequences. The authors introduce a novel feature which called "CSS" based on the self-similarity of low-level features capturing pairwise statistics of specially localized colour distributions. Subsequently, the authors firstly evaluate performance of classification coupling HOG features with CSS; then, to the previous group of features, motion information was added and then computed as a variant of Histogram Of Flows (HOF) algorithm proposed by Dalal et al. [151]. The latter approach consistently improved the detection performance both for static images and video sequences, across the two different datasets. In combination with HOG, these two features outperform the state-of-the-art by up to 20%. In [71], Linear SVM was used to classify and evaluate the performance; then, a variant of AdaBoost algorithm (MLPBoost) was tested also on Caltech dataset.

In [75], Bar-Hillel et al. introduced a new approach for learning part based object detection through feature synthesis in pedestrian detection task. The authors considered different families of features, e.g. HOG, LocalMax or Sift, and for each iteration of their algorithm, a subset of features was used in the generation process, by using pruning strategy as well. In details, the described method consists of an iterative process of feature generation and pruning, in which basic part-based features are developed into a feature hierarchy using operators for part localization, part refining and part combination. Then, feature pruning was performed by using a new features selection algorithm for linear SVM, namely Predictive Feature Selection (PFS), based on weight prediction.

In [73], Park et al. described a multi-resolution model that acts as a deformable part-based model when scoring large instances and a rigid template with scoring small instances. Substantially, the authors demonstrated the necessity to extract features and classify at multi-resolution stages to avoid miss-detection. As in [64], latent SVM were used in the classification step, and the authors demonstrated impressive results on the Caltech Pedestrian benchmark.

In [81], Benenson et al. presented a new pedestrian detector that efficiently handles different scales avoiding the resize of input images; by transferring computation from test time to training time, detection speed was optimized and improved. When processing monocular images, the system provides high quality detections at 50 fps. The authors also proposed a new method for exploiting geometric context extracted from stereo images with high fps by using a CPU+GPU machine.

In [84], Lim et al. proposed a novel approach to both learning and detecting local contour-based representations for mid-level features. The features, called sketch tokens, are learned using supervised mid-level information in the form of hand drawn contours in images. Combining sketch tokens features with the integral image approach proposed by Dollar et al. [67,80] the authors reported a slight improvement in the pedestrian detection approach, with a classification achieved by Random Forest Algorithm.

In [85], Benenson et al. revisited the core assumptions of HOG+SVM algorithm and showed that, by properly designing the feature pooling, feature selection, pre-processing, and training methods, it is possible to reach high performance in pedestrian detection. The authors described an approach based on the multi-scale model generation introduced in [81] but, in contrast with the algorithm proposed by Dalal et al. [56], the windows considered for features extraction selected during learning, were composed by irregular patterns.

In [86], Levi et al. presented a new part-based object detection algorithm with hundreds of parts performing real-time detection based on the approach proposed in [75] by Bar et al.. However, due to their high computational demands part-based methods are limited to several parts only and are too slow for practical real-time implementation. The authors proposed the Accelerated Feature Synthesis (AFS) algorithm and, in order to reduce the number of locations searched for each part, introduced an algorithm for approximate nearest neighbor (KDFerns), to compare each image location to only a subset of the model parts. Candidate part locations for a specific part are then further reduced by using spatial inhibition, and using an object-level "coarse-to-fine" strategy. Linear SVM was used in the classification step.

In [89], Park et al. introduced a combined approach for motion information extraction from video sequences. Prior to features extraction, the authors performed a weak motion stabilization by considering both camera and object motion, and at the same time preserved non-rigid motion that provided useful information for the recognition task. The authors also described a combined approach that used coarse-scale flow and fine-scale temporal difference features and used AdaBoost for classification.

In [93], Zhang et al. proposed a pedestrian detection algorithm introducing several efficient features based on Haar wavelets, called "Compact Features". In the reported work, the authors assume that pedestrians, in the most of cases, show a recurrent behavior, or rather the first visible part of each pedestrian is the upper-right (head and right part of the shoulder). Following this approach, the authors employed a statistical model of the up-right human body where the head, the upper body, and the lower body are treated as separated parts and, in this way, partial occlusions were handled allowing to reach high performances with an occlusion higher than 35%. The classifier used in [93] was AdaBoost.

More recently, Cao et al. proposed a pedestrian detection algorithm considering a set of features appearance constancy and shape symmetry, called NNNF, constituted by both Non-Neighboring (NNF) and Neighboring Features (NF) [152]. The proposed approach have been tested on Caltech dataset reporting good performances compared with state-of-the-art methods.

4.1.1. 2D vs 3D

Features extraction consists in different methods that transform one or more input images into a reduced representation that could be used as input to classifiers. Thanks to different strategies, it is also possible to reduce the dimensionality of these patterns allowing a faster and more accurate classification [153–158]. For the aim of pedestrian detection, the extraction of features is a fundamental task and it is independent from the acquisition technology. In such a kind of applications, pedestrian detection is performed considering RGB images in both 2D and 3D applications. The additional information coming from the third dimension, regardless of stereo-vision or depth cameras, is especially used for tracking pedestrians, allowing to keep track of their position in a 3D space; this kind of approach is used in applications of PPSs where the mutual positions of pedestrians and the moving object (i.e. an autonomous car) are of fundamental importance to correctly control the object.

5. Machine learning techniques for pedestrian detection

Data mining techniques, including machine learning, have been used to learn hidden information in data in order to train automatic systems for decision making processes in several domains [159,160].

Since an acquired scene may contain several kinds of objects candidate for tracking, image processing techniques often fail to filter out background and/or objects of other classes; thus, machine learning methods may help in discriminating pedestrian from other classes of objects in the scene. According to the workflows introduced in Section 1, both traditional approaches and deep learning strategies are used for classification.

Regarding traditional approaches applied to pedestrian detection, the dataset created from the features extracted after the processing of input images influence the design of the classification strategy [161]. In particular, from the input dataset point of view, several algorithms for dataset processing, such as normalization or dimensionality reduction, have been developed, and are applied to improve the classification performance [162–164]. Even from the classifiers point of view, there are several classification algorithms used to perform pedestrian detection, the most of which consist in supervised approach, such as Support Vector Machine (SVM), Artificial Neural Network (ANN), or Boosting algorithms.

Regarding Deep Learning strategies, instead, the design of deep classifiers following the workflow reported in Fig. 3, such as Convolutional Neural Networks, the main task to address is the design of the network topology. In fact, the number of hidden layers strongly influences the network performance in terms of both classification accuracy and execution time. Although a universal strategy to design a good classifier does not exists, the tradeoff between classification performance and training time of the classifier should lead the design of the topology. Specifically, a number of layers too low reduce the training time but the model could too simple for the classification task; on the contrary, a number of layers too high could lead to the classifier overfitting on training data reducing the classifier performance on new data.

In the following sections, the traditional approaches of machine learning will be discussed, analysing the most common architectures employed for pedestrian classification. Furthermore, Deep Learning strategies will be introduced and discussed focusing on deep structures applied on the pedestrian detection task.

5.1. The traditional approaches

As could be seen in Table 1, which reports the performance of the algorithms discussed in Section 4.1, almost all the considered works for pedestrian detection and tracking use simple classifiers, such as Support Vectors Machines (SVMs) or boost families.

Since the pedestrian detection and tracking tasks have an high computational cost, especially in real time applications, very often in literature are presented classifier models with low complexity. Linear SVM and weak decision trees with low depths boosted to speed up the learning phase are the most used, since they can lead to lighter decision processes making the image processing part the most important in the decisional process.

Even if the SVM's design, in terms of complexity, is an automatic procedure for selecting Support Vectors [165], we present the Artificial Neural Networks (ANNs) performances on the mentioned benchmark databases [166–171], as the case of Zhao et al. that developed a stereo-system for pedestrian classification [172].

In [145], Gravila and Munder developed a system, called PROTECTOR, constituted by several processing modules, one of which consists in a neural model that classify pedestrian based on textural features extracted from each video frame.

In fact, thanks to suitable optimization strategies [173–176] it is possible to find the optimal topology for an ANN to classify two or more classes in the best way and, by using a multi-objective algorithm, the topology could be optimized, thus allowing a faster classification in various research topic [177–185].

5.2. The deep learning approaches

Recent researches in Artificial Intelligence (AI) led to the spread of modern techniques of machine learning based in deep structures, as reported in novel and innovative works, [110,111,114,117,118,186–188].

Deep Learning strategies have been used for automatic object detection and images segmentation and classification applications.

The most diffused DL architectures are Convolutional Neural Networks, which are able to classify images into several categories, automatically learning features through convolutional layers that combine multiple non-linear processes.

Since the training of a CNN is very time and computation resources consuming, two different approaches have been found in literature for CNNs: (i) Transfer Learning, that allow to "re-train" a pre-trained model on different categories (e.g. use AlexNet to discriminate among different kind of tumours); (ii) Feature Extractors, as CNNs are constituted by several convolutional layers which create different layer of features representations, it is possible to catch each layer output and use it as input to simpler classifiers, such as SVM or ANN.

Based on Convolutional Neural Networks, these approaches have the ability to learn effective hierarchical feature representations that characterize the typical variations observed in visual data, including images and video, which make them very well-suited for the most of visual classification tasks.

For pedestrian detection, Szarvas et al. used CNN to classify pedestrian in images [189]. The authors compared their approach to classical SVM approach with Haar features obtaining higher levels of accuracy. Then, the CNN was used as features extractor and the computed descriptors were used as input to a Gaussian-SVM classifier and the reported results were increased respect to the CNN approach for classification.

Automatic features extraction in also used in the work by Zhang et al., where the authors used faster r-cnn for pedestrian detection [190]. The developed system was composed of two cascaded sub-systems: the first was deputy to detect candidate regions in the image that could contain a pedestrian; the second sub-system, instead, was a Boosted Forest classifier for the pedestrian classification [191,192].

Recently, Li et al. have used neural features by applying fully convolutional neural networks as features extractors [186]. In details, the authors have tested and compared the performance of AdaBoost classifiers by using input extracted at different levels from the network. The reported results on benchmark datasets are very promising (Log-Average Miss Rate about 20%) if compared with those reported in Table 1.

Since occlusions are one of the most discussed problems in literature [25,26,68,82,93,149,193], deep learning allowed to strengthen the detection of single parts in order to find and correctly classify occluded pedestrian [82,88,115,116].

Ouyang and Wang presented a probabilistic pedestrian detection framework to solve the issue related on the inaccurate scores of part detectors when there are occlusions or large deformations [82]. In this framework, a deformable part-based model was used to obtain the scores of part detectors and the visibilities of parts were modeled as hidden variables. In the proposed work, a discriminative deep model based on Restricted Boltzmann Machine (RBM) building blocks was used for learning the visibility relationship among overlapping parts at multiple layers. Experimental results on benchmark datasets showed the effectiveness of the proposed approach. An improved version of the proposed algorithm is reported in [88] where Ouyang et al. proposed a mutual visibility deep model that jointly estimates the visibility statuses of overlapping pedestrians using Gaussian Mixture Model (GMM). The visibility relationship among pedestrians was learned from the deep model for recognizing co-existing pedestrians. Experimental results showed that the mutual visibility deep model effectively improved the pedestrian detection results. In [116], the main idea is to construct multi-parts detectors that covers several scales of different body parts and automatically choose important parts for occlusion handling. At the training stage, each part detector is learned by CNN fine-tuning approach, using a CNN pre-trained on ImageNet Database [27]. At the testing stage, a shifting handling method within a CNN is designed. This method handles the problem that positive proposal windows usually shift away from their corresponding ground truth bounding boxes. Moreover, the part selection is determined by data and the effectiveness of the part pool can be fully explored.

Human body pose recognition is also a well-suited task for DL approaches [112,113,115,119,194,195]. Human body pose recognition in video is a long-standing problem in computer vision with a wide range of applications. However, body pose recognition remains a challenging problem due to the high dimensionality of the input data and the high variability of possible body poses. As reported in the previous section, traditional computer vision-based approaches are mostly based on appearance cues such as textures, edges, colour histograms, foreground silhouettes or hand-crafted local features (such as histogram of gradients (HOG) [56]) rather than motion-based features. Alternatively, psychophysical experiments have shown that motion is a powerful visual cue capable to extract high-level information, including articulated pose [196]. In particular, a combination of hand-crafted features and DL classifier may be a good approach to estimate human pose. For example, in [112], it is shown that deep learning is able to successfully incorporate both RGB and motion features for the task of human body pose detection in video.

However, to estimate human body pose, deep learning approach to predict a single class label per image has to be supported by a high resolution semantic segmentation output. To reach this result, Oliveira et al. [119] used the so called "up-convolutional networks" [115,194]; in contrast to usual classification, which contracts the high-resolution input to a low-resolution output, this kind of networks can take an abstract, low-resolution input and predicts a high-resolution output, such as a full-size image. To reach this goal, it is possible to refine the architecture of Long et al. [115] and apply it to human body part segmentation to use it different contexts, such as robotics.

Regarding robotics, human body parts segmentation can be a very valuable tool, especially when it can be applied both indoor and outdoor. For persons who cannot move their upper body, some of the most basic actions, such as drinking water, is rendered impossible without assistance. Robots could identify human body parts, such as hands or harms, and interact with them to perform some of these tasks. Other applications, such as learning from demonstration and human robot handover can also benefit from accurate human part segmentation. For a learning-from-demonstration task, one could take advantage of the high level description of human parts, considering each of them as an explicit mapping between the human and joints of the robot for learning control actions. A robot that needs to hand a tool to its human counterpart must be able to detect where the hands are to perform the task. Human body part segmentation has been considered a very challenging task in computer vision due to the wide variability of the body parts' appearance, pose and viewpoint; self-occlusion and clothing, also, represents very difficult problems to handle.

In [113] the pose estimation is formulated as a Deep Neural Network (DNN)-based regression problem towards body joints. A cascade of such DNN regressors which results in high precision pose estimates is presented. The considered approach has the advantage of reasoning about pose in a holistic fashion and has a simple but yet powerful formulation which capitalizes on recent advances in Deep Learning. DNNs have shown outstanding performance on visual classification tasks [27] and more recently on object localization [197,198].

Table 2
Metrics for Log-Average Miss Rate evaluation considering performances on Inria and Caltech datasets.

	Log-Average Miss Rate			
Dataset	*Mean(%)*	*Std*	*Min(%)*	*Max(%)*
Inria	33.33	21.22	13.32	81.70
Caltech	60.94	16.13	34.60	94.73

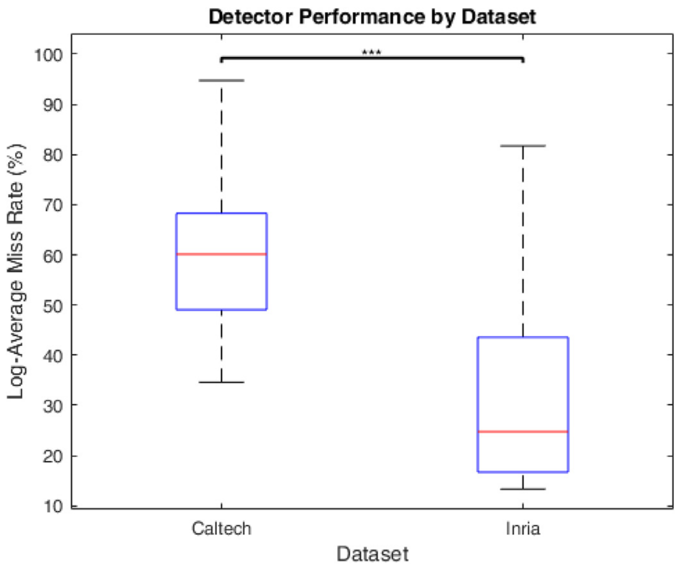

Fig. 9. Box plot of the Log-Average Miss Rate for the different detectors applied on Inria and Caltech Test Sets ($^{*}p \leq 0.05$ $^{**}p \leq 0.01$ $^{***}p \leq 0.001$).

Table 3
Metrics for Log-Average Miss Rate evaluation considering performances of the AdaBoost and SVM classifiers.

	Log-Average Miss Rate			
Algorithm	*Mean(%)*	*Std*	*Min(%)*	*Max(%)*
AdaBoost	44.40	28.21	13.32	94.73
SVM	53.43	16.84	19.96	79.78

6. Discussion and future trends

In the last decades, pedestrian detection and tracking systems gained a considerable importance thanks to their versatility use. The study and development of systems able to automatically interact with moving humans have introduced the need to increase the performance of human detection and, at the same time, improve the run-time performance.

A deep analysis of the results reported in Table 1 is necessary. Table 2 shows different metrics related to the performances of pedestrian detection systems on Inria and Caltech datasets (mean, standard deviation, min and max values). As could be seen, the mean value of Log-Average Miss Rate is significantly higher for detectors on Caltech dataset than on Inria one (lower is better) as shown in Fig. 9 ($p \leq 0.005$). Conversely, the classification performances are not specifically related to the considered classifier family (Fig. 10), even though AdaBoost performs better than SVM on average, as reported in Table 3.

This important result highlights that the implemented classifiers perform better on static images classification than videos containing more noisy classes that could disturb the pedestrian detection. As demonstrated by the works that have shown the best performances on Caltech dataset [89,93], the detection and classification of pedestrian in videos has to be supported by sets of features that take into account motion information too.

A further analysis have been conducted by analysing the performance by grouping classifiers and test sets; in detail, four groups have been created which were: G1 - AdaBoost classifier on Inria; G2 - AdaBoost classifier on Caltech; G3 - SVM classifier on Inria; G4 - SVM classifier on Caltech. The results reported in Table 4 confirm the higher capabilities of classifiers to discriminate pedestrians on the static images from Inria dataset, regardless of the considered classifier families (Fig. 11).

The introduction of Deep Learning architectures, as well as the accessibility of cheaper but more powerful computers, led the scientific community to study more performing systems for two main reasons: (i) DL architectures may help to design more informing sets of features; (ii) DL architectures performance at execution time are faster than traditional models of machine learning.

The computer vision systems adopted to perform pedestrian detection differ based on the acquisition sensor; in particular, 2D sensors limit the task of pedestrian detection to a bi-dimensional space. On the other side, stereo-cameras and depth sensors are able to track pedestrians in the 3D space.

Moreover, some of the works reported in this survey make use of markers; the kind of marker, as well as the aim of each work and the desired level of accuracy to be reach, strongly influence the computer-vision system for images acquisition.

Since human tracking is applied in multiple scenarios, the literature reports a very large variety of configurations for the vision system but, at the same time, the classifiers used to discriminate humans, or pedestrians, among the multitude of objects in the scenes, are limited to the simpler classifiers, such as SVM or decision trees. Artificial Neural Networks are quite used, but very lim-

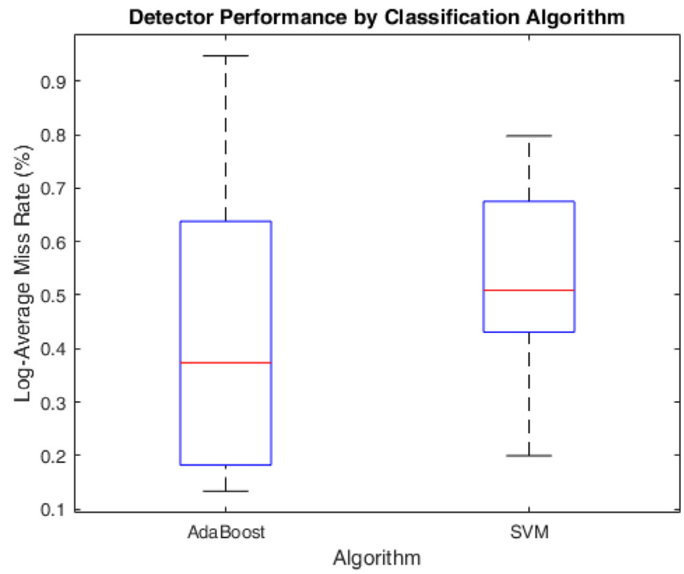

Fig. 10. Box plot of the Log-Average Miss Rate for the different detectors applied on Inria and Caltech Test Sets ($^{*}p \leq 0.05$ $^{**}p \leq 0.01$ $^{***}p \leq 0.001$).

Table 4
Metrics for Log-Average Miss Rate evaluation considering performances on Inria and Caltech datasets.

	Log-Average Miss Rate			
Group	*Mean(%)*	*Std*	*Min(%)*	*Max(%)*
G1	30,81	26,62	13,32	81,70
G2	59,69	22,47	34,60	94,73
G3	37,10	9,93	19,96	45,98
G4	64,32	10,04	48,45	79,78

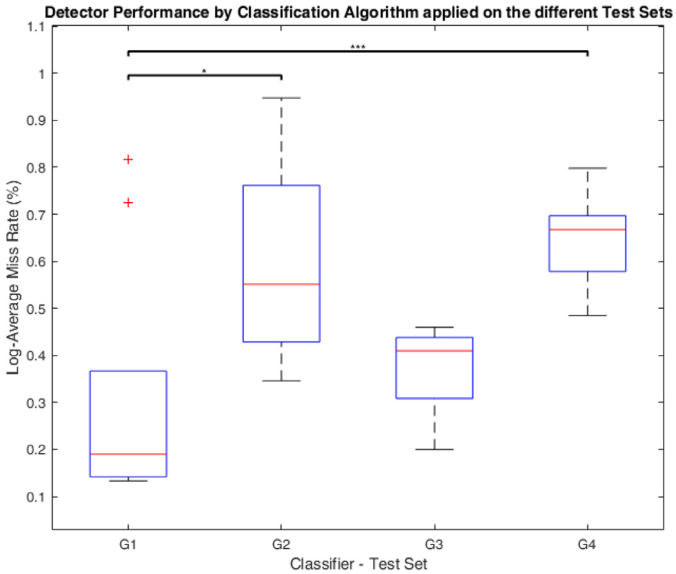

Fig. 11. Box plot of the Log-Average considering four couples of Classifier and Test Set. G1 - AdaBoost classifier on Inria; G2 - AdaBoost classifier on Caltech; G3 - SVM classifier on Inria; G4 - SVM classifier on Caltech ($^{*}p \leq 0.05$ $^{**}p \leq 0.01$ $^{***}p \leq 0.001$).

ited respect to the previous models, besides recent works demonstrated their versatility in different domains [199–202].

The most difficult step in the design of pedestrian detection system concerns with the features extraction, as it is necessary to extract powerful descriptor that have to help to discriminate pedestrian. Thanks to the introduction on Deep Learning structures, the previous step could be bypassed since deep architecture, such as CNNs, could automatically create their own representation of features.

Following the previous conclusions, in order to design novel applications for pedestrian detection, several aspects hat to be considered. First, it is necessary to design the desired degree of accuracy to be reached; this influences the technology of the acquisition system: an RGB camera could be sufficient to detect pedestrian in 2D space, using background subtraction if it is possible crate a simple model of the background; or feature-based approaches to classify pedestrians are to be considered: linear and non-linear classifiers, such as SVM or ANNs, may be considered in cascade to the previous step, or a Deep Learning strategy may be implemented using, for example, Convolutional Neural Networks. In this latter case, there is no need to extract features, but an efficient strategy for objects detection in the image has to be implemented. In any case, a powerful approach may consist in the combination of the two proposed strategies; in details, deep architectures may be used to extract features (even at different levels of abstraction) to be used as input to simple learner for pedestrian classification.

The necessity to track pedestrian in the 3D space imposes the use of depth cameras, or stereo-cameras. This combination is necessary when both the presence and the movements of pedestrian control one or more automatic machine in the real world, such as drones or autonomous vehicles. In some cases, for example in controlled indoor environment, it is necessary "to help" the tracking system with markers. These scenarios are the most common approach in applications that involve drones (even if scientific community is taking the lead of marker-less strategies) or research purposes, such as the study of crowded places tracking and analysing the pedestrian trajectories.

For clinical purposes, instead, the use of multiple markers placed on human body is strongly recommended to accurately track human body parts. In fact, there are only few works that track people without any kind of support for clinical purposes.

6.1. Future works

The future of pedestrian detection concerns with the improvement of performance of both detectors and classifiers. In fact, improving the speed of pedestrian detection has been an active area in recent years. For example, in [81], Benenson et al. proposed a method reaching speeds of 100–135 FPS for detection in a 480×640 image, although the levels of accuracy are still low. Other researchers have focused specifically on speeding up Deep Neural Networks [198,203,204], but with no real-time solutions. In [205], Angelova et al. presented a new real-time approach to object detection that exploits the efficiency of cascade classifiers with the accuracy of deep neural networks.

Excellent performance of Deep Networks in classification tasks are found in literature, and their ability to operate on raw pixel input without the need to design special features is very appealing. However, deep nets are notoriously slow at inference time. In this work, the authors proposed an approach that cascades deep nets and fast features, that is both very fast and very accurate. They applied it to the challenging task of pedestrian detection. Their algorithm runs in real-time at 15 fps. The resulting approach achieves a 26.2% average miss rate on the Caltech Pedestrian detection benchmark, which is competitive with the very best reported results. The importance of pedestrian real-time detection is particularly relevant in advanced driver assistance systems (ADASs), and pedestrian protection systems (PPSs) [76]. As a future work, it could be interesting to find the best trade-off between accuracy and frames per second in different environmental conditions and contexts (e.g. Human-Aware Navigation to detect falls [206]).

7. Conclusion

In this work, a survey on pedestrian detection and tracking system have been presented. Recent adoption of Deep Learning methodologies and in particular of Convolutional Neural Networks for pedestrian detection and tracking deserved a dedicated state-of-the-art survey. The analysed works highlight the need to investigate how modern approaches to pedestrian detection work and a comparison with the features-based approaches on benchmark datasets has to be done.

However, the reported works show encouraging results in automatic pedestrian detection, but further architectures need to be implemented and tested. In particular, for pedestrian detection, the most successful way seems to consist in the combination of Deep Learning with classical Machine Learning models because this seems to imply high levels of accuracy and less computation respect to hand-designed features and classification. Moreover, it will be interesting to compare the performance to this task of optimal ANNs topologies with SVM.

References

[1] C. Urmson, et al., Self-driving cars and the urban challenge, IEEE Intell. Syst. 23 (2) (2008).
[2] M.S. Bartlett, G. Littlewort, I. Fasel, J.R. Movellan, Real time face detection and facial expression recognition: Development and applications to human computer interaction., in: Proceedings of the Conference on Computer Vision and Pattern Recognition Workshop (CVPRW), 5, IEEE, 2003, p. 53.
[3] J. Levinson, J. Askeland, J. Becker, J. Dolson, D. Held, S. Kammel, J.Z. Kolter, D. Langer, O. Pink, V. Pratt, et al., Towards fully autonomous driving: Systems and algorithms, in: Proceedings of the Intelligent Vehicles Symposium (IV), IEEE, 2011, pp. 163–168.
[4] P. Majaranta, A. Bulling, Eye tracking and eye-based human–computer interaction, in: Proceedings of the Advances in Physiological Computing, Springer, 2014, pp. 39–65.

[5] H. Hasan, S. Abdul-Kareem, Human–computer interaction using vision-based hand gesture recognition systems: a survey, Neural Comput. Appl. 25 (2) (2014) 251–261.
[6] M.G. Helander, Handbook of Human-Computer Interaction, Elsevier, 2014.
[7] N. Kos' Myna, F. Tarpin-Bernard, B. Rivet, Bidirectional feedback in motor imagery bcis: learn to control a drone within 5 minutes, in: Proceedings of the Extended Abstracts on Human Factors in Computing Systems (CHI), ACM, 2014, pp. 479–482.
[8] S.S. Rautaray, A. Agrawal, Vision based hand gesture recognition for human computer interaction: a survey, Artif. Intell. Rev. 43 (1) (2015) 1–54.
[9] K. Boudjit, C. Larbes, Detection and implementation autonomous target tracking with a quadrotor ar. drone, in: Proceedings of the 12th International Conference on Informatics in Control, Automation and Robotics (ICINCO), 2, IEEE, 2015, pp. 223–230.
[10] H. Rhodin, N. Robertini, D. Casas, C. Richardt, H.-P. Seidel, C. Theobalt, General automatic human shape and motion capture using volumetric contour cues, in: Proceedings of the European Conference on Computer Vision, Springer, 2016, pp. 509–526.
[11] N. Kardaris, I. Rodomagoulakis, V. Pitsikalis, A. Arvanitakis, P. Maragos, A platform for building new human-computer interface systems that support online automatic recognition of audio-gestural commands, in: Proceedings of the ACM on Multimedia Conference, ACM, 2016, pp. 1169–1173.
[12] V. Bevilacqua, A. Di Maio, A computer vision and control algorithm to follow a human target in a generic environment using a drone, in: D. Huang, K. Han, A. Hussain (Eds.), Proceedings of the 12th International Conference, ICIC Intelligent Computing Methodologies, Lanzhou, China, Part III, Lecture Notes in Computer Science, 9773, Springer, 2016, pp. 192–202, doi:10.1007/978-3-319-42297-8_19.
[13] Y. Imamura, S. Okamoto, J.H. Lee, Human tracking by a multi-rotor drone using hog features and linear SVM on images captured by a monocular camera, in: Proceedings of the International MultiConference of Engineers and Computer Scientists, 1, 2016, pp. 8–13.
[14] N. Unies, U.N.E.C. for Europe, et al., Statistics of Road Traffic Accidents in Europe and North America, New York: United Nations, 2015.
[15] N.H.T.S. Administration, et al., Traffic safety facts, 2012 data: Pedestrians, Ann. Emerg. Med. 65 (4) (2015) 452.
[16] Centers for Disease Control and Prevention, WISQARS (Web-based Injury Statistics Query and Reporting System). Atlanta, GA: US Department of Health and Human Services, CDC, 2015.
[17] L.F. Beck, A.M. Dellinger, M.E. O'neil, Motor vehicle crash injury rates by mode of travel, united states: using exposure-based methods to quantify differences, Am. J. Epidemiol. 166 (2) (2007) 212–218.
[18] T.B. Moeslund, E. Granum, A survey of computer vision-based human motion capture, Comput. Vis. Image Underst. 81 (3) (2001) 231–268.
[19] T.B. Moeslund, A. Hilton, V. Krüger, A survey of advances in vision-based human motion capture and analysis, Comput. Vis. Image Underst. 104 (2) (2006) 90–126.
[20] A. Solichin, A. Harjoko, A.E. Putra, A survey of pedestrian detection in video, Neural Netw. 2 (2014) 41–47.
[21] H. Zhou, H. Hu, Human motion tracking for rehabilitationa survey, Biomed. Signal Process. Control 3 (1) (2008) 1–18.
[22] M. Enzweiler, D.M. Gavrila, Monocular pedestrian detection: Survey and experiments, IEEE Trans. Pattern Anal. Mach. Intell. 31 (12) (2009) 2179–2195.
[23] T. Gandhi, M.M. Trivedi, Pedestrian protection systems: Issues, survey, and challenges, IEEE Trans. Intell. Transp. Syst. 8 (3) (2007) 413–430.
[24] D. Geronimo, A.M. Lopez, A.D. Sappa, T. Graf, Survey of pedestrian detection for advanced driver assistance systems, IEEE Trans. Pattern Anal. Mach. Intell. 32 (7) (2010) 1239–1258, doi:10.1109/TPAMI.2009.122.
[25] P. Dollar, C. Wojek, B. Schiele, P. Perona, Pedestrian detection: An evaluation of the state of the art, IEEE Trans. Pattern Anal. Mach. Intell. 34 (4) (2012) 743–761.
[26] R. Benenson, M. Omran, J. Hosang, B. Schiele, Ten Years of Pedestrian Detection, What Have We Learned?, in: L. Agapito, M. Bronstein, C. Rother (Eds.) Computer Vision - ECCV 2014 Workshops. ECCV 2014, Lecture Notes in Computer Science, vol 8926, Springer, Cham, 2015.
[27] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Proceedings of the Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.
[28] G.E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R.R. Salakhutdinov, Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580, 2012.
[29] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, Caffe: Convolutional architecture for fast feature embedding, in: Proceedings of the 22nd ACM international conference on Multimedia, ACM, 2014, pp. 675–678.
[30] M.D. Zeiler, R. Fergus, Visualizing and Understanding Convolutional Networks, Springer International Publishing, Cham, pp. 818–833. 10.1007/978-3-319-10590-1_53
[31] D. Tomè, F. Monti, L. Baroffio, L. Bondi, M. Tagliasacchi, S. Tubaro, Deep convolutional neural networks for pedestrian detection, Signal Process. Image Commun. 47 (2016) 482–489.
[32] S. Saripalli, J.F. Montgomery, G.S. Sukhatme, Visually guided landing of an unmanned aerial vehicle, IEEE Trans. Robot. Autom. 19 (3) (2003) 371–380.
[33] R. Mobus, U. Kolbe, Multi-target multi-object tracking, sensor fusion of radar and infrared, in: Proceedings of the Intelligent Vehicles Symposium, IEEE, 2004, pp. 732–737.
[34] C.-C. Wang, C. Thorpe, S. Thrun, M. Hebert, H. Durrant-Whyte, Simultaneous localization, mapping and moving object tracking, Int. J. Robot. Res. 26 (9) (2007) 889–916.
[35] J. Sachs, M. Aftanas, S. Crabbe, M. Drutarovsky, R. Klukas, D. Kocur, T.-T. Nguyen, P. Peyerl, J. Rovnakova, E. Zaikov, Detection and tracking of moving or trapped people hidden by obstacles using ultra-wideband pseudo-noise radar, in: Proceedings of the European Radar Conference (EuRAD), IEEE, 2008, pp. 408–411.
[36] S. Gammeter, A. Gassmann, L. Bossard, T. Quack, L. Van Gool, Server-side object recognition and client-side object tracking for mobile augmented reality, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2010, pp. 1–8.
[37] J.-X. Mi, D.-S. Huang, B. Wang, X. Zhu, The nearest-farthest subspace classification for face recognition, Neurocomputing 113 (2013) 241–250.
[38] X.-F. Wang, D.-S. Huang, H. Xu, An efficient local chan–vese model for image segmentation, Pattern Recognit. 43 (3) (2010) 603–618.
[39] L. Carnimeo, V. Bevilacqua, L. Cariello, G. Mastronardi, Retinal vessel extraction by a combined neural network–wavelet enhancement method, Emerg. Intell. Comput. Technol. Appl. Aspects Artif. Intell. (2009) 1106–1116.
[40] X.-F. Wang, D.-S. Huang, A novel multi-layer level set method for image segmentation, J. Univ. Comput. Sci 14 (14) (2008) 2428–2452.
[41] L. Shang, D.-S. Huang, C.-H. Zheng, Z.-L. Sun, Noise removal using a novel non-negative sparse coding shrinkage technique, Neurocomputing 69 (7) (2006) 874–877.
[42] A. Jaimes, N. Sebe, Multimodal human–computer interaction: A survey, Comput. Vis. Image Underst. 108 (1) (2007) 116–134.
[43] V. Bevilacqua, D. Barone, F. Cipriani, G. D'Onghia, G. Mastrandrea, G. Mastronardi, M. Suma, D. D'Ambruoso, A new tool for gestural action recognition to support decisions in emotional framework, in: Proceedings of the IEEE International Symposium on Innovations in Intelligent Systems and Applications (INISTA), IEEE, 2014a, pp. 184–191.
[44] V. Bevilacqua, N. Nuzzolese, D. Barone, M. Pantaleo, M. Suma, D. D'Ambruoso, A. Volpe, C. Loconsole, F. Stroppa, Fall detection in indoor environment with kinect sensor, in: Proceedings of the IEEE International Symposium on Innovations in Intelligent Systems and Applications (INISTA), IEEE, 2014b, pp. 319–324.
[45] M. Billinghurst, A. Clark, G. Lee, et al., A survey of augmented reality, Found. Trends Hum. Comput. Interact. 8 (2–3) (2015) 73–272.
[46] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, B. MacIntyre, Recent advances in augmented reality, IEEE Comput. Graph. Appl. 21 (6) (2001) 34–47, doi:10.1109/38.963459.
[47] C. Kamath, S. Cheung, Robust techniques for background subtraction in urban traffic video, Technical Report, Lawrence Livermore National Laboratory (LLNL), Livermore, CA, 2003.
[48] M. Piccardi, Background subtraction techniques: a review, in: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, 4, IEEE, 2004, pp. 3099–3104.
[49] N.A. Mandellos, I. Keramitsoglou, C.T. Kiranoudis, A background subtraction algorithm for detecting and tracking vehicles, Expert Syst. Appl. 38 (3) (2011) 1619–1631.
[50] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, T. Poggio, Pedestrian detection using wavelet templates, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, 1997, pp. 193–199.
[51] I. Haritaoglu, D. Harwood, L.S. Davis, W/sup 4: real-time surveillance of people and their activities, IEEE Transactions on pattern analysis and machine intelligence 22 (8) (2000) 809–830.
[52] R. Cutler, L.S. Davis, Robust real-time periodic motion detection, analysis, and applications, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (8) (2000) 781–796.
[53] S.L. Dockstader, A.M. Tekalp, Multiple camera tracking of interacting and occluded human motion, Proceedings of the IEEE 89 (10) (2001) 1441–1455.
[54] P. Viola, M.J. Jones, D. Snow, Detecting pedestrians using patterns of motion and appearance, in: Proceedings Ninth IEEE International Conference on Computer Vision, IEEE, 2003, pp. 734–741vol.2, doi:10.1109/ICCV.2003.1238422.
[55] L. Wang, T. Tan, H. Ning, W. Hu, Silhouette analysis-based gait recognition for human identification, IEEE transactions on pattern analysis and machine intelligence 25 (12) (2003) 1505–1518.
[56] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 1, IEEE, 2005, pp. 886–893.
[57] Z.-L. Sun, D.-S. Huang, Y.-M. Cheun, Extracting nonlinear features for multispectral images by FCMC and KPCA, Digital Signal Process. 15 (4) (2005a) 331–346.
[58] Z.-L. Sun, D.-S. Huang, Y.-M. Cheung, J. Liu, G.-B. Huang, Using fcmc, fvs, and pca techniques for feature extraction of multispectral images, IEEE Geosci. Remote Sens. Lett. 2 (2) (2005b) 108–112.
[59] S. Munder, D.M. Gavrila, An experimental study on pedestrian classification, IEEE Trans. Pattern Anal. Mach. Intell. 28 (11) (2006) 1863–1868.
[60] P. Sabzmeydani, G. Mori, Detecting pedestrians by learning shapelet features, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2007, pp. 1–8.
[61] L. Zhang, S.Z. Li, X. Yuan, S. Xiang, Real-time object classification in video surveillance based on appearance learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2007, pp. 1–8.

[62] D. Ramanan, D.A. Forsyth, A. Zisserman, Tracking people by learning their appearance, IEEE transactions on pattern analysis and machine intelligence 29 (1) (2007) 65–81.
[63] S. Maji, A.C. Berg, J. Malik, Classification using intersection kernel support vector machines is efficient, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2008, pp. 1–8.
[64] P. Felzenszwalb, D. McAllester, D. Ramanan, A discriminatively trained, multiscale, deformable part model, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2008, pp. 1–8.
[65] Y. Lu, S. Payandeh, Cooperative hybrid multi-camera tracking for people surveillance, Canad. J. Electr. Comput. Eng. 33 (3/4) (2008) 145–152.
[66] B. Li, D.-S. Huang, C. Wang, K.-H. Liu, Feature extraction using constrained maximum variance mapping, Pattern Recognit. 41 (11) (2008) 3287–3294.
[67] P. Dollár, Z. Tu, P. Perona, and S. Belongie, Integral Channel Features, Proc. British Machine Vision Conf., 2009.
[68] X. Wang, T.X. Han, S. Yan, An HOG-LBP human detector with partial occlusion handling, in: Proceedings of the IEEE 12th International Conference on Computer Vision, IEEE, 2009, pp. 32–39.
[69] M.W. Lee, R. Nevatia, Human pose tracking in monocular sequence using multilevel structured models, IEEE Trans. Pattern Anal. Mach. Intell. 31 (1) (2009) 27–38.
[70] B. Li, C. Wang, D.-S. Huang, Supervised feature extraction based on orthogonal discriminant projection, Neurocomputing 73 (1) (2009) 191–196.
[71] S. Walk, N. Majer, K. Schindler, B. Schiele, New features and insights for pedestrian detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2010, pp. 1030–1037.
[72] P.F. Felzenszwalb, R.B. Girshick, D. McAllester, D. Ramanan, Object detection with discriminatively trained part-based models, IEEE Trans. Pattern Anal. Mach. Intell. 32 (9) (2010) 1627–1645.
[73] D. Park, D. Ramanan, C. Fowlkes, Multiresolution models for object detection, in: Proceedings of the Computer Vision–ECCV, 2010, pp. 241–254.
[74] O. Oreifej, R. Mehran, M. Shah, Human identity recognition in aerial images, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2010, pp. 709–716.
[75] A. Bar-Hillel, D. Levi, E. Krupka, C. Goldberg, Part-based feature synthesis for human detection, in: Proceedings of the Computer Vision–ECCV, 2010, pp. 127–142.
[76] D. Geronimo, A.M. Lopez, A.D. Sappa, T. Graf, Survey of pedestrian detection for advanced driver assistance systems, IEEE Trans. Pattern Anal. Mach. Intell. 32 (7) (2010) 1239–1258.
[77] B. Benfold, I. Reid, Stable multi-target tracking in real-time surveillance video, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2011, pp. 3457–3464.
[78] P.D.Z. Varcheie, G.-A. Bilodeau, Adaptive fuzzy particle filter tracker for a PTZ camera in an IP surveillance system, IEEE Trans. Instrum. Measur. 60 (2) (2011) 354–371.
[79] J. Sherrah, B. Ristic, N. Redding, Particle filter to track multiple people for visual surveillance, IET Comput. Vis. 5 (4) (2011) 192–200.
[80] P. Dollár, R. Appel, W. Kienzle, Crosstalk cascades for frame-rate pedestrian detection, in: Proceedings of the Computer Vision–ECCV, Springer, 2012, pp. 645–659.
[81] R. Benenson, M. Mathias, R. Timofte, L. Van Gool, Pedestrian detection at 100 frames per second, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2012, pp. 2903–2910.
[82] W. Ouyang, X. Wang, A discriminative deep model for pedestrian detection with occlusion handling, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2012, pp. 3258–3265.
[83] Y. Zhao, D.-S. Huang, W. Jia, Completed local binary count for rotation invariant texture classification, IEEE Trans. Image Proces. 21 (10) (2012) 4492–4497.
[84] J.J. Lim, C.L. Zitnick, P. Dollár, Sketch tokens: A learned mid-level representation for contour and object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3158–3165.
[85] R. Benenson, M. Mathias, T. Tuytelaars, L. Van Gool, Seeking the strongest rigid detector, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3666–3673.
[86] D. Levi, S. Silberstein, A. Bar-Hillel, Fast multiple-part based object detection using KD-ferns, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 947–954.
[87] J. Yan, X. Zhang, Z. Lei, S. Liao, S.Z. Li, Robust multi-resolution pedestrian detection in traffic scenes, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3033–3040.
[88] W. Ouyang, X. Zeng, X. Wang, Modeling mutual visibility relationship in pedestrian detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3222–3229.
[89] D. Park, C.L. Zitnick, D. Ramanan, P. Dollár, Exploring weak stabilization for motion feature extraction, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2013, pp. 2882–2889.
[90] T. Naseer, J. Sturm, D. Cremers, Followme: Person following and gesture recognition with a quadrocopter, in: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2013, pp. 624–630.
[91] J. Portmann, S. Lynen, M. Chli, R. Siegwart, People detection and tracking from aerial thermal views, in: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2014, pp. 1794–1800.
[92] H. Cho, Y.-W. Seo, B.V. Kumar, R.R. Rajkumar, A multi-sensor fusion system for moving object detection and tracking in urban driving environments, in: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA),, IEEE, 2014, pp. 1836–1843.
[93] S. Zhang, C. Bauckhage, A.B. Cremers, Informed haar-like features improve pedestrian detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 947–954.
[94] F. De Smedt, D. Hulens, T. Goedemé, On-board real-time tracking of pedestrians on a UAV, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2015, pp. 1–8.
[95] Z.-P. Bian, J. Hou, L.-P. Chau, N. Magnenat-Thalmann, Fall detection based on body part tracking using a depth camera, IEEE J. Biomed. Health Inf. 19 (2) (2015) 430–439.
[96] F. Mueller, M. Muirhead, Jogging with a quadcopter, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, ACM, 2015, pp. 2023–2032.
[97] K.K. Lekkala, V.K. Mittal, Simultaneous aerial vehicle localization and human tracking, in: Proceedings of the Region 10 Conference (TENCON), IEEE, 2016, pp. 379–383.
[98] Y. Ma, X. Wu, G. Yu, Y. Xu, Y. Wang, Pedestrian detection and tracking from low-resolution unmanned aerial vehicle thermal imagery, Sensors 16 (4) (2016) 446.
[99] W.G. Aguilar, M.A. Luna, J.F. Moya, V. Abad, H. Parra, H. Ruiz, Pedestrian detection for uavs using cascade classifiers with meanshift, in: Proceedings of the IEEE 11th International Conference on Semantic Computing (ICSC), IEEE, 2017, pp. 509–514.
[100] M. Pediaditis, M. Tsiknakis, N. Leitgeb, Vision-based motion detection, analysis and recognition of epileptic seizuresa systematic review, Comput. Methods Programs Biomed. 108 (3) (2012) 1133–1148.
[101] I.F. Mondragón, P. Campoy, M.A. Olivares-Mendez, C. Martinez, 3d object following based on visual information for unmanned aerial vehicles, in: Proceedings of the IEEE IX Latin American Robotics Symposium and IEEE Colombian Conference on Automatic Control and Industry Applications (LARC), IEEE, 2011, pp. 1–7.
[102] C. Martínez, I.F. Mondragón, P. Campoy, J.L. Sánchez-López, M.A. Olivares-Méndez, A hierarchical tracking strategy for vision-based applications on-board UAVS, J. Intell. Robot. Syst. 72 (3–4) (2013) 517–539.
[103] F. Guérin, F. Guinand, J.-F. Brethé, H. Pelvillain, A. Zentout, et al., Vision based target tracking using an unmanned aerial vehicle, in: Proceedings of the IEEE International Workshop on Advanced Robotics and its Social Impacts (ARSO), IEEE, 2015, pp. 1–6.
[104] W. Mehner, M. Boltes, M. Mathias, B. Leibe, Robust marker-based tracking for measuring crowd dynamics, in: Proceedings of the International Conference on Computer Vision Systems, Springer, 2015, pp. 445–455.
[105] F. Vasconcelos, N. Vasconcelos, Person-following UAVS, in: Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2016, pp. 1–9.
[106] I. Bortone, G.F. Trotta, A. Brunetti, G.D. Cascarano, C. Loconsole, N. Agnello, A. Argentiero, G. Nicolardi, A. Frisoli, V. Bevilacqua, A novel approach in combination of 3d gait analysis data for aiding clinical decision-making in patients with Parkinson's disease, in: Proceedings of the International Conference on Intelligent Computing, Springer, Cham, 2017, pp. 504–514.
[107] V. Bevilacqua, G.F. Trotta, A. Brunetti, N. Caporusso, C. Loconsole, G.D. Cascarano, F. Catino, P. Cozzoli, G. Delfine, A. Mastronardi, et al., A comprehensive approach for physical rehabilitation assessment in multiple sclerosis patients based on gait analysis, in: Proceedings of the International Conference on Applied Human Factors and Ergonomics, Springer, Cham, 2017a, pp. 119–128.
[108] V. Bevilacqua, G.F. Trotta, C. Loconsole, A. Brunetti, N. Caporusso, G.M. Bellantuono, I. De Feudis, D. Patruno, D. De Marco, A. Venneri, et al., A rgb-d sensor based tool for assessment and rating of movement disorders, in: Proceedings of the International Conference on Applied Human Factors and Ergonomics, Springer, Cham, 2017b, pp. 110–118.
[109] V.M. Manghisi, A.E. Uva, M. Fiorentino, V. Bevilacqua, G.F. Trotta, G. Monno, Real time RULA assessment using Kinect v2 sensor, Appl. Ergon. 65 (2017) 481–491.
[110] S. Ji, W. Xu, M. Yang, K. Yu, 3d convolutional neural networks for human action recognition, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 221–231.
[111] N. Wang, D.-Y. Yeung, Learning a deep compact image representation for visual tracking, in: Proceedings of the Advances in Neural Information Processing Systems, 2013, pp. 809–817.
[112] A. Jain, J. Tompson, Y. LeCun, C. Bregler, Modeep: A deep learning framework using motion features for human pose estimation, in: Proceedings of the Asian Conference on Computer Vision, Springer, 2014, pp. 302–315.
[113] A. Toshev, C. Szegedy, Deeppose: Human pose estimation via deep neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1653–1660.
[114] L. Wang, W. Ouyang, X. Wang, H. Lu, Visual tracking with fully convolutional networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 3119–3127.
[115] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.
[116] Y. Tian, P. Luo, X. Wang, X. Tang, Deep learning strong parts for pedestrian detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1904–1912.
[117] H. Xue, Y. Liu, D. Cai, X. He, Tracking people in rgbd videos using deep learning and motion clues, Neurocomputing 204 (2016) 70–76.

[118] H. Nam, B. Han, Learning multi-domain convolutional neural networks for visual tracking, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4293–4302.
[119] G.L. Oliveira, A. Valada, C. Bollen, W. Burgard, T. Brox, Deep learning for human part discovery in images, in: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2016, pp. 1634–1641.
[120] B.L. Sefidgari, Feed-back method based on image processing for detecting human body via flying robot, Int. J. Artif. Intell. Appl. 4 (6) (2013) 35.
[121] K. Pfeil, S.L. Koh, J. LaViola, Exploring 3d gesture metaphors for interaction with unmanned aerial vehicles, in: Proceedings of the International Conference on Intelligent User Interfaces, ACM, 2013, pp. 257–266.
[122] M. Kimura, R. Shibasaki, X. Shao, M. Nagai, Automatic extraction of moving objects from UAV-borne monocular images using multi-view geometric constraints, in: Proceedings of the IMAV International Micro Air Vehicle Conference and Competition, Delft University of Technology, Delft, The Netherlands, 2014.
[123] J. Nagi, A. Giusti, G.A. Di Caro, L.M. Gambardella, Human control of uavs using face pose estimates and hand gestures, in: Proceedings of the ACM/IEEE International Conference on Human-robot interaction, ACM, 2014, pp. 252–253.
[124] K. Miyoshi, R. Konomura, K. Hori, Above your hand: direct and natural interaction with aerial robot, in: ACM SIGGRAPH 2014 Emerging Technologies, ACM, 2014, p. 8.
[125] C. Pittman, J.J. LaViola Jr, Exploring head tracked head mounted displays for first person robot teleoperation, in: Proceedings of the 19th International Conference on Intelligent User Interfaces, ACM, 2014, pp. 323–328.
[126] F. Mueller, M. Muirhead, Understanding the design of a flying jogging companion, in: Proceedings of the Adjunct Publication of the 27th Annual ACM Symposium on User Interface Software and Technology, ACM, 2014, pp. 81–82.
[127] D. Buongiorno, M. Barsotti, E. Sotgiu, C. Loconsole, M. Solazzi, V. Bevilacqua, A. Frisoli, A neuromusculoskeletal model of the human upper limb for a myoelectric exoskeleton control using a reduced number of muscles, in: Proceedings of the IEEE World Haptics Conference (WHC), 2015, pp. 273–279, doi:10.1109/WHC.2015.7177725.
[128] D. Buongiorno, F. Barone, M. Solazzi, V. Bevilacqua, A. Frisoli, A linear optimization procedure for an emg-driven neuromusculoskeletal model parameters adjusting: Validation through a myoelectric exoskeleton control, in: Proceedings of the International Conference on Human Haptic Sensing and Touch Enabled Computer Applications, Springer, 2016, pp. 218–227.
[129] D. Buongiorno, F. Barone, D.J. Berger, B. Cesqui, V. Bevilacqua, A. dAvella, A. Frisoli, Evaluation of a pose-shared synergy-based isometric model for hand force estimation: Towards myocontrol, in: Proceedings of the Converging Clinical and Engineering Research on Neurorehabilitation II, Springer, 2017, pp. 953–958.
[130] V.I. Morariu, D. Harwood, L.S. Davis, Tracking people's hands and feet using mixed network and/or search, IEEE Trans. Pattern Anal. Mach. Intell. 35 (5) (2013) 1248–1262.
[131] S. Yao, S. Pan, T. Wang, C. Zheng, W. Shen, Y. Chong, A new pedestrian detection method based on combined hog and lss features, Neurocomputing 151 (2015) 1006–1014, doi:10.1016/j.neucom.2014.08.080.
[132] G. Lu, Y. Yan, L. Ren, P. Saponaro, N. Sebe, C. Kambhamettu, Where am i in the dark: Exploring active transfer learning on the use of indoor localization based on thermal imaging, Neurocomputing 173 (2016) 83–92, doi:10.1016/j.neucom.2015.07.106.
[133] B. Sheng, Q. Hu, J. Li, W. Yang, B. Zhang, C. Sun, Filtered shallow-deep feature channels for pedestrian detection, Neurocomputing 249 (2017) 19–27, doi:10.1016/j.neucom.2017.03.007.
[134] C. Zhu, Y. Peng, Discriminative latent semantic feature learning for pedestrian detection, Neurocomputing 238 (2017) 126–138, doi:10.1016/j.neucom.2017.01.043.
[135] M. Boltes, A. Seyfried, Collecting pedestrian trajectories, Neurocomputing 100 (2013) 127–133.
[136] M. Munaro, E. Menegatti, Fast RGB-D people tracking for service robots, Autonom. Robots 37 (3) (2014) 227–242.
[137] H. Sun, C. Wang, B. Wang, N. El-Sheimy, Pyramid binary pattern features for real-time pedestrian detection from infrared videos, Neurocomputing 74 (5) (2011) 797–804, doi:10.1016/j.neucom.2010.10.009.
[138] Y. Xia, W. Xu, L. Zhang, X. Shi, K. Mao, Integrating 3d structure into traffic scene understanding with RGB-D data, Neurocomputing 151 (2015) 700–709, doi:10.1016/j.neucom.2014.05.091.
[139] L. Maddalena, A. Petrosino, A self-organizing approach to background subtraction for visual surveillance applications, IEEE Trans. Image Process. 17 (7) (2008) 1168–1177.
[140] S. Brutzer, B. Höferlin, G. Heidemann, Evaluation of background subtraction techniques for video surveillance, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2011, pp. 1937–1944.
[141] Y. Benezeth, P.-M. Jodoin, B. Emile, H. Laurent, C. Rosenberger, Review and evaluation of commonly-implemented background subtraction algorithms, in: Pattern Recognition, 2008. ICPR 2008. 19th International Conference on, IEEE, 2008, pp. 1–4.
[142] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. Pattern Anal. Mach. Intell. 11 (7) (1989) 674–693.
[143] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 1, IEEE, 2001. I–I
[144] P. Viola, M.J. Jones, Robust real-time face detection, International journal of computer vision 57 (2) (2004) 137–154.
[145] D.M. Gavrila, S. Munder, Multi-cue pedestrian detection and tracking from a moving vehicle, Int. J. Comput. Vis. 73 (1) (2007) 41–59.
[146] S. Munder, C. Schnorr, D.M. Gavrila, Pedestrian detection and tracking using a mixture of view-based shape–texture models, IEEE Trans. Intell. Transp. Syst. 9 (2) (2008) 333–343.
[147] T. Ojala, M. Pietikainen, D. Harwood, Performance evaluation of texture measures with classification based on kullback discrimination of distributions, in: Proceedings of the 12th IAPR International Conference on Pattern Recognition Computer Vision & Image Processing, 1, IEEE, 1994, pp. 582–585.
[148] P. Dollár, C. Wojek, B. Schiele, P. Perona, Pedestrian detection: A benchmark, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2009, pp. 304–311.
[149] Y. Cheng, Mean shift, mode seeking, and clustering, IEEE Trans. Pattern Anal. Mach. Intell. 17 (8) (1995) 790–799, doi:10.1109/34.400568.
[150] J.P. Lewis, Fast template matching, in: Vision Interface, 95, 1995, pp. 15–19.
[151] N. Dalal, B. Triggs, C. Schmid, Human detection using oriented histograms of flow and appearance, in: Proceedings of the European Conference on Computer Vision, Springer, 2006, pp. 428–441.
[152] J. Cao, Y. Pang, X. Li, Pedestrian detection inspired by appearance constancy and shape symmetry, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1316–1324.
[153] C.-H. Zheng, D.-S. Huang, Z.-L. Sun, M.R. Lyu, T.-M. Lok, Nonnegative independent component analysis based on minimizing mutual information technique, Neurocomputing 69 (7) (2006) 878–883.
[154] Z.-L. Sun, D.-S. Huang, C.-H. Zheng, L. Shang, Optimal selection of time lags for TDSEP based on genetic algorithm, Neurocomputing 69 (7) (2006) 884–887.
[155] Z.-Q. Zhao, D.-S. Huang, W. Jia, Palmprint recognition with 2DPCA+ PCA based on modular neural networks, Neurocomputing 71 (1) (2007) 448–454.
[156] D.-S. Huang, J.-X. Mi, A new constrained independent component analysis method, IEEE Trans. Neural Netw. 18 (5) (2007) 1532–1535.
[157] B. Li, C.-H. Zheng, D.-S. Huang, Locally linear discriminant embedding: An efficient method for face recognition, Pattern Recognit. 41 (12) (2008) 3813–3821.
[158] X.-F. Wang, D.-S. Huang, A novel density-based clustering framework by using level set method, IEEE Trans. Knowl. Data Eng. 21 (11) (2009) 1515–1531.
[159] D. Huang, The Study of Data Mining Methods for Gene Expression Profiles, Science Press of China, 2009.
[160] F. Menolascina, S. Tommasi, A. Paradiso, M. Cortellino, V. Bevilacqua, G. Mastronardi, Novel data mining techniques in ACGH based breast cancer subtypes profiling: the biological perspective, in: Proceedings of the IEEE Symposium on Computational Intelligence and Bioinformatics and Computational Biology, IEEE, 2007, pp. 9–16.
[161] A. Janecek, W. Gansterer, M. Demel, G. Ecker, On the relationship between feature selection and classification accuracy, in: New Challenges for Feature Selection in Data Mining and Knowledge Discovery, 2008, pp. 90–105.
[162] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504–507.
[163] H.-L. Wei, S.A. Billings, Feature subset selection and ranking for data dimensionality reduction, IEEE Trans. Pattern Anal. Mach. Intell. 29 (1) (2007).
[164] C.-I. Chang, Hyperspectral Data Processing: Algorithm Design and Analysis, Wiley, USA, NJ, Englewood, 2013.
[165] M. Rychetsky, Algorithms and architectures for machine learning based on regularized neural networks and support vector approaches, Shaker Verlag GmbH, 2001, pp. 3–36.
[166] D.-S. Huang, in: Systematic Theory of Neural Networks for Pattern Recognition, 201, Publishing House of Electronic Industry of China, Beijing, 1996.
[167] D.-s. Huang, Radial basis probabilistic neural networks: Model and application, Int. J. Pattern Recognit. Artif. Intell. 13 (07) (1999) 1083–1101.
[168] D.-S. Huang, J.-X. Du, A constructive hybrid structure optimization methodology for radial basis probabilistic neural networks, IEEE Trans. Neural Netw. 19 (12) (2008) 2099–2115.
[169] J.-X. Du, D.-S. Huang, G.-J. Zhang, Z.-F. Wang, A novel full structure optimization algorithm for radial basis probabilistic neural networks, Neurocomputing 70 (1) (2006) 592–596.
[170] D.-S. Huang, S.-D. Ma, Linear and nonlinear feedforward neural network classifiers: a comprehensive understanding, J. Intell. Syst. 9 (1) (1999) 1–38.
[171] D.-S. Huang, W.-B. Zhao, Determining the centers of radial basis probabilistic neural networks by recursive orthogonal least square algorithms, Appl. Math. Comput. 162 (1) (2005) 461–473.
[172] L. Zhao, C.E. Thorpe, Stereo-and neural network-based pedestrian detection, IEEE Trans. Intell. Transp. Syst. 1 (3) (2000) 148–154.
[173] V. Bevilacqua, G. Mastronardi, F. Menolascina, P. Pannarale, A. Pedone, A novel multi-objective genetic algorithm approach to artificial neural network topology optimisation: the breast cancer classification problem, in: Proceedings of the International Joint Conference on Neural Networks(IJCNN), IEEE, 2006, pp. 1958–1965.
[174] V. Bevilacqua, A. Brunetti, M. Triggiani, D. Magaletti, M. Telegrafo, M. Moschetta, An optimized feed-forward artificial neural network topology to support radiologists in breast lesions classification, in: Proceedings of the on Genetic and Evolutionary Computation Conference Companion, ACM, 2016, pp. 1385–1392.
[175] W.-B. Zhao, D.-S. Huang, J.-Y. Du, L.-M. Wang, Genetic optimization of radial

basis probabilistic neural networks, Int. J. Pattern Recognit. Artif. Intell. 18 (08) (2004) 1473–1499.
[176] D.-S. Huang, W. Jiang, A general CPL-ADS methodology for fixing dynamic parameters in dual environments, IEEE Trans. Syst. Man Cybern. Part B Cybern. 42 (5) (2012) 1489–1500.
[177] V. Bevilacqua, N. Pietroleonardo, V. Triggiani, A. Brunetti, A.M. Di Palma, M. Rossini, L. Gesualdo, An innovative neural network framework to classify blood vessels and tubules based on haralick features evaluated in histological images of kidney biopsy, Neurocomputing 228 (2017a) 143–153.
[178] V. Bevilacqua, S. Simeone, A. Brunetti, C. Loconsole, G.F. Trotta, S. Tramacere, A. Argentieri, F. Ragni, G. Criscenti, A. Fornaro, et al., A computer aided ophthalmic diagnosis system based on tomographic features, in: Proceedings of the International Conference on Intelligent Computing, Springer, Cham, 2017b, pp. 598–609.
[179] V. Bevilacqua, A. Brunetti, G.F. Trotta, G. Dimauro, K. Elez, V. Alberotanza, A. Scardapane, A novel approach for hepatocellular carcinoma detection and classification based on triphasic ct protocol, in: Proceedings of the IEEE Congress on Evolutionary Computation (CEC), IEEE, 2017c, pp. 1856–1863.
[180] W. Jiang, D.-S. Huang, S. Li, Random walk-based solution to triple level stochastic point location problem, IEEE Trans. Cybern. 46 (6) (2016) 1438–1451.
[181] C.-Y. Lu, D.-S. Huang, Optimized projections for sparse representation based classification, Neurocomputing 113 (2013) 213–219.
[182] Z.-Q. Zhao, D.-S. Huang, A mended hybrid learning algorithm for radial basis function neural networks to improve generalization capability, Appl. Math. Modell. 31 (7) (2007) 1271–1281.
[183] L. Shang, D.-S. Huang, J.-X. Du, C.-H. Zheng, Palmprint recognition using fastica algorithm and radial basis probabilistic neural network, Neurocomputing 69 (13) (2006) 1782–1786.
[184] D.-S. Huang, H.H.-S. Ip, K.C.K. Law, Z. Chi, Zeroing polynomials using modified constrained neural network approach, IEEE Trans. Neural Netw. 16 (3) (2005) 721–732.
[185] D.-S. Huang, A constructive approach for finding arbitrary roots of polynomials by neural networks, IEEE Trans. Neural Netw. 15 (2) (2004) 477–491.
[186] C. Li, X. Wang, W. Liu, Neural features for pedestrian detection, Neurocomputing 238 (2017) 420–432, doi:10.1016/j.neucom.2017.01.084.
[187] V. Bevilacqua, D. Altini, M. Bruni, M. Riezzo, A. Brunetti, C. Loconsole, A. Guerriero, G.F. Trotta, R. Fasano, M. Di Pirchio, et al., A supervised breast lesion images classification from tomosynthesis technique, in: Proceedings of the International Conference on Intelligent Computing, Springer, Cham, 2017, pp. 483–489.
[188] M. Raza, C. Zonghai, S.U. Rehman, W. Peng, W. Ji-kai, Framework for estimating distance and dimension attributes of pedestrians in real-time environments using monocular camera, Neurocomputing (2017), doi:10.1016/j.neucom.2017.08.052.
[189] M. Szarvas, A. Yoshizawa, M. Yamamoto, J. Ogata, Pedestrian detection with convolutional neural networks, in: Proceedings of the Intelligent Vehicles Symposium, IEEE, 2005, pp. 224–229.
[190] L. Zhang, L. Lin, X. Liang, K. He, Is faster r-cnn doing well for pedestrian detection? in: Proceedings of the European Conference on Computer Vision, Springer, 2016, pp. 443–457.
[191] J. Friedman, T. Hastie, R. Tibshirani, et al., Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors), Ann. Stat. 28 (2) (2000) 337–407.
[192] R. Appel, T. Fuchs, P. Dollár, P. Perona, Quickly boosting decision trees–pruning underachieving features early, in: Proceedings of the 30th International Conference on Machine Learning (ICML-13), 2013, pp. 594–602.
[193] X. Zhang, H.-M. Hu, F. Jiang, B. Li, Pedestrian detection based on hierarchical co-occurrence model for occlusion handling, Neurocomputing 168 (2015) 861–870, doi:10.1016/j.neucom.2015.05.038.
[194] A. Dosovitskiy, J. Tobias Springenberg, T. Brox, Learning to generate chairs with convolutional neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1538–1546.
[195] M. Raza, Z. Chen, S.-U. Rehman, P. Wang, P. Bao, Appearance based pedestrians' head pose and body orientation estimation using deep learning, Neurocomputing (2017), doi:10.1016/j.neucom.2017.07.029.
[196] G. Johansson, Visual perception of biological motion and a model for its analysis, Percept. Psychophys. 14 (2) (1973) 201–211.
[197] C. Szegedy, A. Toshev, D. Erhan, Deep neural networks for object detection, in: Proceedings of the Advances in Neural Information Processing Systems, 2013, pp. 2553–2561.
[198] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 580–587.
[199] J.-X. Du, D.-S. Huang, X.-F. Wang, X. Gu, Computer-aided plant species identification (CAPSI) based on leaf shape matching technique, Trans. Inst. Measur. Control 28 (3) (2006) 275–285.
[200] J.-X. Du, D.-S. Huang, X.-F. Wang, X. Gu, Shape recognition based on neural networks trained by differential evolution algorithm, Neurocomputing 70 (4) (2007) 896–903.
[201] X.-F. Wang, D.-S. Huang, J.-X. Du, H. Xu, L. Heutte, Classification of plant leaf images with complicated background, Appl. Math. Comput. 205 (2) (2008) 916–926.
[202] V. Bevilacqua, L. Carnimeo, G. Mastronardi, V. Santarcangelo, R. Scaramuzzi, On the comparison of nn-based architectures for diabetic damage detection in retinal images, J. Circ. Syst. Comput. 18 (08) (2009) 1369–1380.
[203] A. Giusti, D.C. Ciresan, J. Masci, L.M. Gambardella, J. Schmidhuber, Fast image scanning with deep max-pooling convolutional neural networks, in: Proceedings of the 20th IEEE International Conference on Image Processing (ICIP), IEEE, 2013, pp. 4034–4038.
[204] F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, K. Keutzer, Densenet: implementing efficient convnet descriptor pyramids, arXiv preprint arXiv:1404.1869(2014).
[205] A. Angelova, A. Krizhevsky, V. Vanhoucke, A.S. Ogale, D. Ferguson, Real-time pedestrian detection with deep network cascades., in: Proceedings of the BMVC, 2015. 32–1
[206] D. Ribeiro, A. Mateus, J.C. Nascimento, P. Miraldo, A real-time pedestrian detector using deep learning for human-aware navigation, arXiv preprint arXiv:1607.04441(2016).

Antonio Brunetti received the Bachelor Degree in Computer Science and Automation Engineering, and the Master Degree (cum laude) in Computer Science Engineering at the Polytechnic University of Bari. His final dissertation in Human-Computer Interaction concerned the design and implementation of an intelligent system able to adapt the living environment of subjects with mild cognitive impairment based on the cognitive event-related potential on the EEG signal. Subsequently, he worked as Research Fellow at the Polytechnic University of Bari working on GCESYS – "Green Community Efficiency Systems" project dealing with optimization algorithms applied to energy consumptions scheduling in smart buildings with systems supplied by mixed traditional and renewable sources. Currently, he is a PhD student in Electrical and Information Engineering at the Doctoral School of the Polytechnic University of Bari under the supervision of Prof. Vitoantonio Bevilacqua, working on Decision Support Systems based on biomedical signals for the customization and optimization of diagnosis, prognosis and therapy. Some of his professional experience includes several collaborations at the Department of Electrical and Information Engineering (DEI) of the Polytechnic University of Bari where he worked on the design and development of intelligent algorithms for image and signal processing, working on different projects, such as RESCAP – "Residual Capabilities", and PRE.C.I.O.U.S. – "PREdictive Computer aIded scOring sUpport System". Currently, he is also a collaborator on PRE.MED – "PREdictive MEDicine" project, where he has to design and develop an operating environment for the extraction and processing of morphological features of patients affected by Autosomal Dominant Polycystic Kidney Disease (ADPKD) and to validate the diagnostic classification framework.

Domenico Buongiorno received the B.Sc. and M.Sc. (cum laude) degrees in automation and control theory engineering from Polytechnic University of Bari, Italy, in 2011 and 2014, respectively. In 2017, he received the Ph.D. degree in Emerging Digital Technologies from Scuola Superiore Sant'Anna, Pisa, Italy (PERCeptual RObotics laboratory – PERCRO) with a topic in the control of robotic interfaces for interaction with virtual environments, robot-aided neurorehabilitation (EMG based control – Myocontrol) and bilateral multi-DoF teleoperation in disaster scenario. In 2017, he was a visiting PhD Student at the Biorobotics Laboratory, UCI Irvine, California. He is currently a Post-Doc Researcher at the Department of Electrical and Information Engineering of Polytechnic University of Bari and his research interests include human-machine and human-robot interaction, assistive technologies, muscle synergy analysis and myocontrol for neuro-rehabilitation.

Gianpaolo Francesco Trotta received the Bachelor Degree in Computer Science and Automation Engineering, and the Master Degree in Computer Science Engineering at the Polytechnic University of Bari. His final dissertation in Human-Computer Interaction concerned the design and implementation of the interactions of subjects with mild cognitive impairment in a virtual scenario. Currently, he is a PhD student in Mechanical and Management Engineering at the Doctoral School of the Polytechnic University of Bari under the supervision of Prof. Vitoantonio Bevilacqua, Prof. Antonio Emanuele Uva and Prof. Michele Fiorentino, working on the design and the implementation of interactions in virtual, augmented and mixed reality in medical field applications and the design and the implementation of systems able to objectify measures during medical examinations by means of low-cost devices. From February to April 2018, he is a visiting PhD Student at the Institute of Computer Graphics and Vision, Technology University of Graz, Austria, under the supervision of the Prof. Jan Egger. Some of his professional experience includes several collaborations at the Department of Electrical and Information Engineering (DEI) of the Polytechnic University of Bari where he worked on Study and Implementation of Information Extraction Algorithms of Editorial Interest from Multimedia Information Content, working on SKIN – "System of Knowledge delivered by an Innovative Network" project and Study and implementation of a web platform for the management and processing of vital parameters and biomarkers, working on PRE.C.I.O.U.S. – "PREdictive Computer aIded scOring sUpport System" project.

Vitoantonio Bevilacqua obtained the Laurea Degree in Electronic Engineering and the Ph.D. in Electrical Engineering from Polytechnic University of Bari, was qualified as Associate Professor of Information Processing Systems and in March 2017 also as Associate Professor of Bioengineering. Currently he is Professor of Human Computer Interaction at the Department of Electrical and Information Engineering of Polytechnic University of Bari where previously also taught Expert Systems, Medical Informatics and Image Processing, and where is the Head of Industrial Informatics Lab. Since 1996 he has been working and investigating in the field of computer vision and image processing, human-machine interaction, bioengineering, machine learning and soft computing (neural networks, evolutionary algorithms, hybrid expert systems, deep learning). The main applications of his research are in medicine, in biometry, in bioinformatics in ambient assisted living and industry. In 2000 he was involved as Visiting Researcher in an EC funded TMR (Trans-Mobility of Researchers) network (ERB FMRX-CT97-0127) called CAMERA (CAd Modeling Environment from Range Images) and worked in Manchester (UK) at UK Robotics Ltd, in the field of geometric feature extraction and 3D objects reconstruction. He has published more than 145 papers in refereed journals, books, international conferences proceedings and chaired several sessions such as Speech Recognition, Biomedical Informatics, Intelligent Image Processing and Bioinformatics in international conferences. On July 2011, he was invited as lecturer at International School on Medical Imaging using Bioinspired and Soft Computing-Miere (Spain) MIBISOC FP7-PEOPLE-ITN-2008. GA N. 238819 where presented his research on Intelligent Tumors Computer Aided Early Diagnosis and Therapy: Neural Network and Genetic Algorithms frameworks. He served as 2009 and 2016 International Conference on Intelligent Computing Program Chair, as 2015 and 2017 Human Machine Interaction Summer School General Chair and as ISMISS 2016 IEEE Italy Section Medical Informatics Summer School Co-Chair.