

Pedestrian Detection in Far Infrared Images based on the use of Probabilistic Templates

M. Bertozzi¹ A. Broggi¹ C. Hilario Gomez² R. I. Fedriga¹ G. Vezzi¹ M. Del Rose³

Abstract—This article presents a validator stage for a pedestrian detection system based on the use of probabilistic models for the infrared domain. Four different models are employed in order to recognize the pose of the pedestrians; open, almost open, almost closed and fully closed legs are detected. In an attempt to overcome the drawbacks of template-matching in far infrared images, two different approaches are proposed. The algorithm has been tested on an experimental vehicle in different situations and a Receiver Operating Characteristic has been computed.

I. INTRODUCTION

In the last years, interests and topics of research on Intelligent Transportation Systems have evolved. Recently, an increasing focus has been paid on pedestrian protection, in order to avoid the high number of casualties or injuries. Several public and private research groups and therefore actively pursuing and developing research in pedestrian detection systems. In fact, the automatic recognition of people or obstacles can improve safety both for regular vehicles and for vulnerable road users.

In addition, this technology could be greatly beneficial for a broad number of fields, such as driver assistance systems, surveillance systems, and intelligent systems for autonomous or semi-autonomous driving. In particular, the U.S. Army is actively developing obstacle detection for mule operations, path following and intent based anti-tamper surveillance or its robotic vehicles safety [11], [14].

Unfortunately, pedestrian detection is a challenging task and the complexity of the problem is augmented by factors like the use of moving cameras, uncontrolled outdoor environments and variations in pedestrian's appearance and pose. In order to solve this problem, different approaches have been chosen and tested.

This work presents a module for pedestrian validation and classification that works on infrared images. It is part of a more complex system that combines the output from a visible and a far infrared stereo systems, with the aim of exploiting the benefits of both approaches. The system exploits the use of probabilistic templates for human shape detection.

This paper is organized as follows: section II describes related work in pedestrian detection based on artificial vision. The pedestrian detection system is discussed in III and the module for pedestrian validation and classification is

detailed in section IV. Finally, section V and VI present the experimental results and conclusions.

II. RELATED WORK

For the U. S. Army the use of vision as a primary sensor for the detection of human shapes is a natural choice since cameras are non-evasive sensors.

Vision-based systems for pedestrian detection have been developed exploiting different approaches; like the use of monocular [21], [24], [30] or stereo [9], [18], [22], [25] vision. Many systems based on the use of a stationary camera employ simple segmentation techniques to obtain foreground region; but this approach fails when the pedestrians has to be detected from moving platforms. Most of the current approaches for pedestrian detection using moving cameras treats the problem as a recognition task: a foreground detection is followed by a recognition step to verify the presence of a pedestrian. Some systems use motion detection [9], [20] or stereo analysis [31] as a means of segmentation.

Other systems, substitute the segmentation step with a focus-of-attention approach, where salient regions in feature maps are considered as candidates for pedestrians. In the GOLD system [3], vertical symmetries are associated with potential pedestrians. In [8] the local image entropy directs the focus-of-attention followed by a model-matching module.

For what concerns the recognition phase, recent research are often motion-based, shape-based or multi-cue based. Motion based approaches use the periodicity of human gait [8], [9], [15], [17], [28] or gait patterns for pedestrian detection [28]. These approaches seems to be more reliable than shape based ones, but they require temporal information and are unable to correctly classify pedestrians that are still or have an unusual gait pattern.

On the other hand, shape-based approaches are more sensible to false positives, so their performance rely on the goodness of the detection phase. These methods are based on pedestrians' appearance, so can detect both moving and stationary people [3], [5], [12]. In these cases, the challenge is to model the several variations in the shapes, pose, size and appearance of humans and their backgrounds. In [13] a tree-based hierarchy of human silhouettes is constructed and the matching follows a coarse-to-fine approach. In [16], [23] probabilistic templates are used to take in account the possible variations in human shape. To accomplish the recognition step, other systems systems exploit pattern recognition based on the use of classifiers [13], [19] or in combination with a shape analysis with gait detection [18]. Most advanced systems rely on histograms of oriented gradient descriptors

¹VisLab – Dip. Ing. Informazione, Università degli Studi di Parma, ITALY, www.vislab.it

²Intelligent Systems Lab, Universidad Carlos III de Madrid, SPAIN, www.uc3m.es/islab

³U.S.Army TARDEC, Warren, MI, U.S.A., www.tacom.army.mil

for human detection, employing SVM for the recognition step [10], or they built detectors based on both motion and appearance information, using AdaBoost for the training process [27].

Recently, infrared technologies (both far and near infrared based) are becoming a popular approach for pedestrian detection [2], [29] due to the decreasing cost of infrared technology. Far infrared (FIR) cameras have demonstrated to be adequate when the background is colder than the human shape [1], in low illumination conditions or during the night [26]. FIR images generally are less sensitive to noise, easing the detection process, but these approaches can fail if the background is warmer than the pedestrians or if in the scene there are other objects that emit heat. Moreover, human body does not emit heat uniformly. Therefore infrared is not a complete solution to the task of pedestrian detection.

III. THE SYSTEM DESCRIPTION

In this paper, a pedestrian detection method based on probabilistic models is proposed. The algorithm has been developed as a part of a tetra-vision based pedestrian detection system [4], [6]; the whole system exploits the simultaneous use of two infrared and visible camera stereo pairs. The main idea is to have both the advantages of far infrared and visible cameras trying at the same time to cope with the deficiencies of each system.

Images in the infrared (IR) domain convey a type of information that is very different from images in the visible spectrum. While in the visible spectrum the image of an object depends on the amount of incident light on its surface and on how well the surface reflects it, in the IR domain the image of an object is related to its temperature and the amount of heat it emits (see Fig. 1).



Fig. 1. Examples of typical FIR scenarios.

Pedestrians usually emit more heat than static objects, such as trees, roads, etc. Therefore image regions containing pedestrians or other "hot" objects are usually brighter than the background. On the other hand, infrared images are blurred, have a poorer resolution and the contrast is low compared with rich and colorful visible images.

Consequently, both visible and far infrared images are used for reducing the search space. Different approaches have been developed for pedestrian detection in the two image domains: warm areas detection, vertical edges detection and an approach based on the simultaneous computation of disparity space images in the two domains. In the FIR domain, a stereo match is performed in order to compute

the size and position in the real world of the bounding boxes detected in the previous steps.

These first stages of detection output a list of bounding boxes that enclose potential pedestrians. Since different approaches are used in both domains to process the same scene, different bounding boxes often belong to the same obstacle; therefore, a merging process is mandatory. A symmetry-based approach is furtherly used to refine this rough result. Then, a number of validators are used to evaluate the presence of a human shape inside each bounding box. The validation is performed searching for human shape characteristics: head detection, shape detection and an active contour-based approach. Figure 2 sketches the overall algorithm flow.

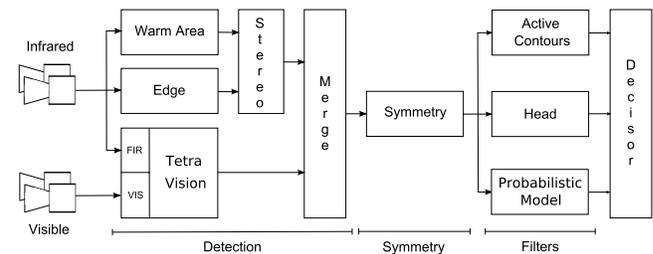


Fig. 2. Overall algorithm flow.

IV. THE PROBABILISTIC APPROACH

The overall shape of the object is also evaluated by means of probabilistic models. The algorithm returns to the system the probability that a box contains a human shape. This probability is combined with the other decisors' response to take a final decision about the presence of a pedestrian.

In the following the validation process based on the use of match against probabilistic models is detailed.

A. Feature Extraction for Infrared Images

The performance of a pedestrian detection system depends heavily on the features extracted. In far infrared images, the intensity of the pixels are representative of the temperature of the objects. This can ease the detection phase, when pedestrians are expected to appear with a brighter intensity than the background. Unfortunately, other objects in the infrared image produce regions that might be as bright or brighter than pedestrians' shapes. Therefore, the recognition task can not depend on high intensity value alone. As an alternative, pattern recognition techniques can be used to reliably classify humans.

In addition, the intensity value in an infrared image depends on other factors besides the temperature, such as the body part, type of clothes and pose of the person. For instance, it is typical that pedestrian regions concerning the torso area are darker than head-regions and hand-regions. As a result, different body parts have different amount of variations in their intensity. Therefore, uniform template-matching or segmentation into regions are barely effective. Besides, traditional edge-based approaches are of no use due to the low contrast and ghosting effects that may appear with high temperature conditions.

Consequently, a pixel-based representation is used. The raw intensity values at each pixel are used to classify target objects between pedestrians or non-pedestrians.

B. Human Shape Models' creation

In the following, the way of creating the models is explained. A training data set of 240 images has been collected for developing the probabilistic templates (see figure 3). Pedestrians walking in front or laterally with respect to the camera axis are considered. Under this scenario, the cyclic nature of the walking action ensures that the shape of each pedestrian is repeated. Consequently, the training data set is previously divided in different classes, depending on the position of the legs during the walking process as: open, almost open, almost closed, and fully closed legs. Each training set contains 60 images of pedestrians with a height of 210 pixels. The width is not selected, since it changes during the walking phase and, in addition, depends on the position of the arms.



Fig. 3. A few examples of training data set images.

For developing the probabilistic templates, firstly, simple thresholding is applied to the images on all training set images.

Determining a threshold that correctly separates every human from its background is not easy. In a previous work [7], the threshold was selected based on heuristics. Anyway, a more precise approach has been developed.

Several images containing only background has been analyzed in order to establish the threshold more precisely. The mean and standard deviation for each pixel in the images was calculated. Because the non-pedestrian class follows a Gaussian distribution, the threshold that differentiates pedestrian from non-pedestrian could be calculated based on the following statement,

$$T = \mu + 3\sigma \quad (1)$$

Since pedestrian's pixels are brighter than background's ones, only the upper limit is considered, ensuring that the 99% of the background pixels is below that threshold. The so obtained threshold is also used for creating the probabilistic models.

Once the threshold is computed, training images are resized and overlapped in a manner that the center of all the images is in the same position. After this normalization step, the probabilistic models are computed: each pixel of the model represents how frequently that pixel is above the threshold namely the probability that it is belonging to a pedestrian. The size of the final models is 210×90 pixels.

Fig. 4 shows the resulting models used for the recognition step.

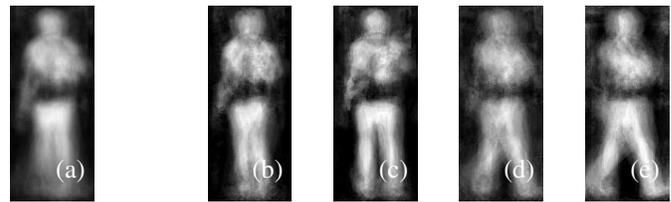


Fig. 4. (a) Mask-model obtained considering all kind of poses and models used for the recognition step: (b) closed (c) almost closed (d) almost open (e) open legs.

The bounding boxes enclosing potential pedestrians feature different sizes depending on the object distance from the two stereovision systems; therefore in order to take in account this the probabilistic models are resampled according to the estimated distance of the potential pedestrian. In the figure 7 is shown how the algorithm can recognize pedestrians located both close and far from the cameras. Since the scaling operations are time expensive, models at different resolutions are computed off-line. This limits the precision of the match, since not all possible resolutions are computed but, at the same time, improve the efficiency of the system.

C. Detection of target objects

Two different methods for extracting target objects from the infrared images have been developed.

In the method proposed in [4] an average mask model (see Fig. 4.a) was employed for performing the background subtraction. To enhance the foreground and remove the background, the following formula was used:

$$th(x, y) = \begin{cases} 255 & \text{if } p(x, y) \geq 0.75 \text{ AND } I(x, y) > 127 \\ 0 & \text{if } p(x, y) \leq 0.20 \text{ AND } I(x, y) \leq 127 \\ I(x, y) & \text{otherwise} \end{cases} \quad (2)$$

Empirically, it was found that regions where $p(x, y) \geq 0.75$ corresponds to head and torso and $p(x, y)$ is lower than 0.20 in correspondence to the background. Intermediate situations are generally due to legs or arms. In this way the foreground is enhanced and consequently the contrast with the background is increased. Figure 5.c shows the results after applying this technique to the input images.

As an alternative to this method, another way to extract target objects has been developed. In this case, objects are segmented using intensity thresholding as expressed in following equation:

$$th(x, y) = \begin{cases} 255 & \text{if } I(x, y) > \text{threshold} \\ 0 & \text{if } I(x, y) \leq \text{threshold} \end{cases} \quad (3)$$

where $th(x, y)$ is the thresholded image, $I(x, y)$ is the raw input image and the value of *threshold* is obtained as explained in (1). The pixels that belong to the background are given a negative value and the pixels that correspond to

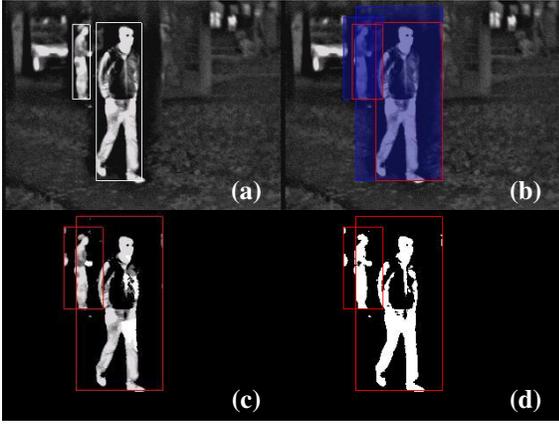


Fig. 5. Foreground enhancement and background removal: (a) original bounding boxes, (b) search areas (blue) and rectangles (red) that correspond to the size of the mask model, (c) result of the first and (d) second approaches.

the foreground are replaced by 1. Fig. 5.d shows the results of this method.

D. Pedestrian detection

Once the foreground is extracted, the recognition phase follows a model matching approach. The four probabilistic models are matched against the extracted potential pedestrians. As mentioned in section III, this algorithm is part of a more complex system and is fed with a list of bounding boxes that contain potential pedestrians to be validated [7].

Since sometimes the bounding boxes are not sufficiently accurate, the search region is enlarged proportionally to the bounding box size (see figure 5.b). In order to get rid of false positives, a filter relying on size and aspect ratio has been devised. After enlarging the bounding boxes, small or huge boxes are discarded, assuming that they can not contain human shapes. After this filtering phase, a list of regions that can contain potential pedestrians is obtained. The probabilistic models are scaled to fit each region and the corresponding matching values are computed.

On this purpose, two different formulas have been tested, being the second one an attempt to cope with the limitations of the first one. In the case in which potential regions were extracted using (2) formula, a correlation $C_1(x, y)$ is computed as:

$$C_1(x, y) = \sum_{i=1}^m \sum_{j=1}^n (th_{xy}(i, j) - 127) * (p(i, j) - 0.5) \quad (4)$$

where m and n are the model width and height and $th_{xy}(i, j)$ is the image after the enhancement operation. The function $C_1(x, y)$ encodes the probability that the area $m \times n$ centered in (x, y) contains a pedestrian. The contribution of a pixel (i, j) in the extracted region th_{xy} is given by $(th_{xy}(i, j) - 127) * (p(i, j) - 0.5)$ (where p are the values of the mask pixels). If a pixel's value $th_{xy}(i, j)$ is higher than 127, it will have a positive contribution proportional to the product of its closeness to white i.e. $(th_{xy}(i, j) - 127)$ and

the confidence measure of it being white i.e. $(p(i, j) - 0.5)$. The same formula works for pixels lower than 127. The contribution is going to be positive when the considered pixel features a value close to the value of model's corresponding pixel (i.e. a black pixel, $th_{xy}(i, j) < 127$) is the same as what it is expected $p < 0.5$.

However, the formula, originally proposed by [16], had an inconvenience when applied to infrared images. Experimental results showed that the intensity of the majority of the pixels was below the given threshold, therefore the contribution to $C_1(x, y)$ of *dark* pixels is higher than the one obtained from *bright* pixels. As a result, most of the pixels give a negative contribution. Moreover, even if the correlation values are within a range, the interval limits are not known.

For these reasons, an improvement was made to this method proposing a new matching formula to face the problem. The matching is performed onto the regions extracted by the second method (3). This time, the mask model used before is not needed. The aim of using such a model, was to enhance the intensity value of the foreground pixels. As a result the contrast with the background was increased and the segmentation task eased. On the contrary, using the equation (3) to extract the target objects, there is no need to increase the intensity of the image, as their values are either 0 (background) or 255 (foreground).

The new matching formula is defined as:

$$C_2(x, y) = \frac{\sum_{i=1}^m \sum_{j=1}^n (th_{xy}(i, j) - 127) \times (p(i, j) - 0.5)}{\sum_{i=1}^m \sum_{j=1}^n |p(i, j) - 0.5|} \quad (5)$$

where $th_{xy}(i, j)$ is the thresholded input image after applying the filter sketched by equation (3). In this way, the contribution of the pixels is proportional to the correspondent values on the probabilistic template; by only considering the sign of the input pixels –positive if it is foreground and negative otherwise– instead of the intensity values, the resulting correlation value is more accurate, as it does not promote some pixels rather than others.

Besides, the matching is calculated separately for the background and the foreground pixels, in an attempt to give the same weight to the contribution of either region. Consequently, the overall correlation value is obtained summing up the two normalized terms,

$$C_2(x, y) = \frac{1}{2} (C_{2\text{background}}(x, y) + C_{2\text{foreground}}(x, y)) \quad (6)$$

In addition, the second formula (5) is normalized between [-1, 1], allowing an easier way to compare the matching values than before applying equation (4). Since four different models are used, the classification in both cases is solved as a maximization problem. The model that gets the higher probability value is considered as the best match and the probability itself is considered as the final vote and fed to the

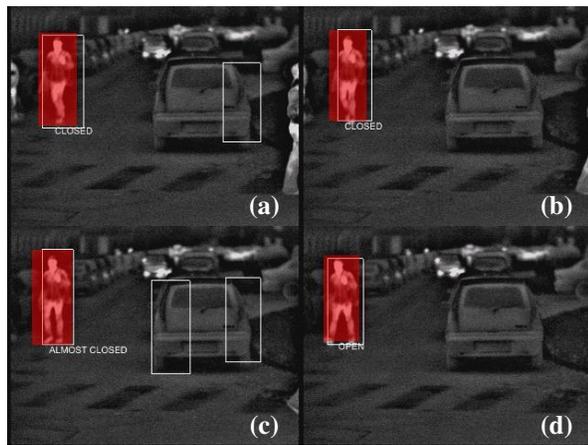


Fig. 6. Pedestrian is labelled as having closed legs in (a) or (b), having almost closed legs in (c) and as open legs in (d).

decisor (see figure 2). Bounding boxes featuring a probability lower than a given threshold are discarded.

Moreover, pose recognition can be achieved because four different models are created considering different position of the legs. Therefore, the model that gets the best matching value is used to label the legs' pose as open, almost open, almost closed, or closed (see Fig. 6).

V. RESULTS

The developed system has been tested in different situations using an experimental vehicle equipped with the tetra-vision system. The proposed algorithm has been run in real time on several FIR sequences.

Figure 7 shows some results of the match with the probabilistic model. If the confidence on the detection is above a given threshold, the bounding box is drawn using red, meaning the region has a high probability of containing a human shape, or using blue if the confidence is low. It can be seen in results that the algorithm is able to detect one or more pedestrians even if they are close to each other, in presence of complex background or partially occluded.

Nonetheless, a template-matching approach fails when the human shape in the image strongly differs from the training set. In figure 7.* pedestrian sitting on a motorcycle is not

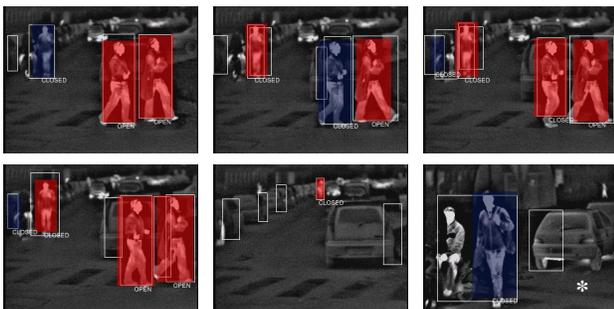


Fig. 7. Results of the probabilistic model validators: detected pedestrians are shown using a superimposed red box when the confidence on the classification is high, blue when the confidence is low, or none for discarded boxes.

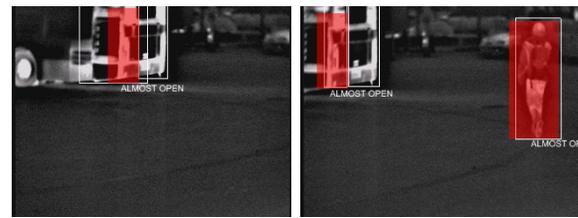


Fig. 8. Wrong results: objects having pedestrian size can be misdected.

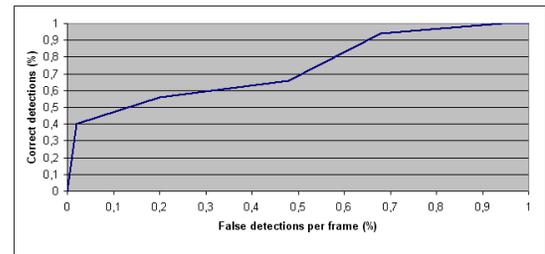


Fig. 9. ROC graph obtained using equation (4).

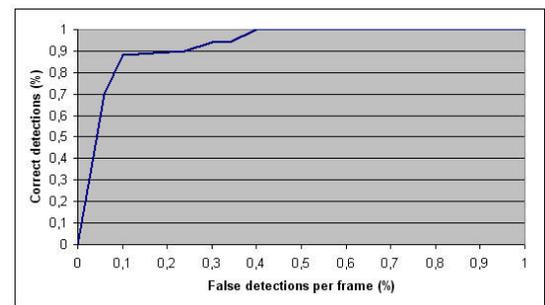


Fig. 10. ROC graph obtained using equation (5).

recognized, since the probabilistic models do not encode humans that are sitting.

The most critical problems concern aspect ratio and imprecise bounding boxes. Sometimes aspect ratio is not a good evaluation criterion for filtering results. Figure 8 shows that some objects in the scene can produce a bounding box that is compatible with a pedestrian. This problem can cause the system to get false positive results despite of the human shape matching filtering.

Another problem concern bounding boxes containing part of a pedestrian. This can cause false negatives, as the human shape is not complete and the matching fails. Similar failures occur due to occlusions or missing parts of the body, but this problem is observed only in few frames, thus tracking could be used to cope with these particular cases. On the other hand, the integration of other validators in the whole system reduces the false positives.

The performance of the algorithm has been evaluated obtaining ROC curves shown in figures 9 and 10 and acting on correlation threshold. The first curve has been obtained running the system using the matching equation (4). Using this method, correct detection percentage smoothly increases. The second curve has been obtained employing the formula (5) onto the regions extracted using equation (3).

The data set used for the ROC curves is composed of 100 bounding boxes containing pedestrian and non-pedestrian. It can be seen that in this condition the system is able to correctly detect more than 85% of pedestrians in scene, with a very low number of false detection per frame. Therefore, the robustness of the algorithm is increased by means of the improvements made to the first implementation.

VI. CONCLUSIONS

In this paper a human shape detection system in FIR images has been presented. The discussed algorithm is based on the use of a match against probabilistic human shape templates and is part of a more complex system [7].

The match-based process has been tested in urban and extraurban environments using an experimental vehicle equipped with two FIR cameras and two daylight cameras.

The algorithm is based on the validation of a list of areas of attention obtained from the tetra-vision system. In order to validate the presence of pedestrians and to detect their poses, four different probabilistic models of human shape are matched against the potential pedestrians.

Experimental results show that the system is promising; the detection rate is high while maintaining a low false positive rate. The system has proven to work also when pedestrians are partially occluded or when parts of the body are missing. False detections are due to the lack of the corresponding model or to the noise produce by other objects in scene.

ACKNOWLEDGMENT

This work has been supported by the European Research Office of the U. S. Army under contract number N62558-05-P-0380.

Ms. Hilario thanks the fellowship awarded by the University Carlos III of Madrid for her stay at Parma.

REFERENCES

- [1] M. Bertozzi, A. Broggi, M. Del Rose, and A. Lasagni. Infrared Stereo Vision-based Human Shape Detection. In *Procs. IEEE Intelligent Vehicles Symposium 2005*, pages 23–28, Las Vegas, USA, June 2005. doi:10.1109/IVS.2005.1505072.
- [2] M. Bertozzi, A. Broggi, A. Fascioli, T. Graf, and M.-M. Meinecke. Pedestrian Detection for Driver Assistance Using Multiresolution Infrared Vision. *IEEE Trans. on Vehicular Technology*, 53(6):1666–1678, Nov. 2004, doi:10.1109/TVT.2004.834878. ISSN 0018-9545.
- [3] M. Bertozzi, A. Broggi, A. Fascioli, and M. Sechi. Shape-based Pedestrian Detection. In *Procs. IEEE Intelligent Vehicles Symposium 2000*, pages 215–220, Detroit, USA, Oct. 2000. doi:10.1109/IVS.2000.898344.
- [4] M. Bertozzi, A. Broggi, M. Felisa, G. Vezzoni, and M. Del Rose. Low-level Pedestrian Detection by means of Visible and Far Infra-red Tetra-vision. In *Procs. IEEE Intelligent Vehicles Symposium 2006*, pages 231–236, Tokyo, Japan, June 2006. doi:10.1109/IVS.2006.1689633.
- [5] D. Beymer and K. Konolige. Real-time Tracking of Multiple People using Continuous Detection. In *Procs. Intl. Conf. on Computer Vision*, Kerkyra, 1999.
- [6] E. Binelli, A. Broggi, A. Fascioli, S. Ghidoni, P. Grisleri, T. Graf, and M.-M. Meinecke. A Modular Tracking System for Far Infrared Pedestrian Recognition. In *Procs. IEEE Intelligent Vehicles Symposium 2005*, pages 758–763, Las Vegas, USA, June 2005.
- [7] A. Broggi, M. Bertozzi, M. Felisa, P. Grisleri, S. Ghidoni, G. Vezzoni, C. Hilario Gómez, and M. Del Rose. Pedestrian Detection by means of Far-infrared Stereo Vision. *Computer Vision and Image Understanding*, 2007, doi:10.1016/j.cviu.2006.07.016.
- [8] C. Curio, J. Edelbrunner, T. Kalinke, C. Tzomakas, and W. von Seelen. Walking Pedestrian Recognition. *IEEE Trans. on Intelligent Transportation Systems*, 1(3):155–163, Sept. 2000.
- [9] R. Cutler and L. S. Davis. Robust Real-time Periodic Motion Detection, Analysis and Applications. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(8):781–796, Aug. 2000.
- [10] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Intl. Conf. on Computer Vision & Pattern Recognition*, volume 2, pages 886–893, June 2005, doi:10.1109/CVPR.2005.177.
- [11] M. Del Rose and P. Frederick. Pedestrian Detection. In *Procs. Intelligent Vehicle Systems Symposium*, Traverse City, USA, 2005.
- [12] H. Fujiyoshi and A. Lipton. Real-time Human Motion Analysis by Image Skeletonization. In *Procs. of the Workshop on Application of Computer Vision*, pages 15–21, Oct. 1998.
- [13] D. M. Gavrilu. Pedestrian Detection from a Moving Vehicle. In *Procs. of European Conference on Computer Vision*, volume 2, pages 37–49, June–July 2000.
- [14] R. Kania, M. Del Rose, and P. Frederick. Autonomous Robotic Following Using Vision Based Techniques. In *Procs. Ground Vehicle Survivability Symposium*, Monterey, USA, 2005.
- [15] H. Mori, N. M. Charkari, and T. Matsushita. On Line Vehicle and Pedestrian Detection based on Sign Pattern. *IEEE Trans. on Industrial Electronics*, 41(4):384–391, Aug. 1994.
- [16] H. Nanda and L. Davis. Probabilistic Template Based Pedestrian Detection in Infrared Videos. In *Procs. IEEE Intelligent Vehicles Symposium 2002*, Paris, France, June 2002.
- [17] A. Niyogi, Sourabh and H. Adelson, Edward. Analyzing and Recognizing Walking Figures in XYT. In *Procs. Conf. on Computer Vision and Pattern Recognition*, pages 469–474, Seattle, USA, June 1994.
- [18] C. Papageorgiou, T. Evgeniou, and T. Poggio. A Trainable Pedestrian Detection System. In *Procs. IEEE Intelligent Vehicles Symposium '98*, pages 241–246, Stuttgart, Germany, Oct. 1998.
- [19] V. Philomin, R. Duraiswami, and L. Davis. Pedestrian Tracking from a Moving Vehicle. In *Procs. IEEE Intelligent Vehicles Symposium 2000*, pages 350–355, Detroit, USA, Oct. 2000.
- [20] R. Polana and R. C. Nelson. Detection and Recognition of Periodic, Non-rigid Motion. *Internation Journal of Computer Vision*, 23(3):261–282, June–July 1997.
- [21] A. Shashua, Y. Gdalyahu, and G. Hayun. Pedestrian Detection for Driving Assistance Systems: Single-frame Classification and System level Performance. In *Procs. IEEE Intelligent Vehicles Symposium 2004*, Parma, Italy, June 2004.
- [22] H. Shimizu and T. Poggio. Direction Estimation of Pedestrian from Multiple Still Images. In *Procs. IEEE Intelligent Vehicles Symposium 2004*, Parma, Italy, June 2004.
- [23] C. Stauffer and W. E. L. Grimson. Similarity Templates for Detection and Recognition. In *Procs. IEEE Intl. Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 221–228, 2001.
- [24] G. P. Stein, O. Mano, and A. Shashua. Vision based ACC with a Single Camera: Bounds on Range and Range Rate Accuracy. In *Procs. IEEE Intelligent Vehicles Symposium 2003*, Columbus, USA, June 2003.
- [25] S. Tate and Y. Takefuji. Video-based Human Shape Detection Deformable Templates and Neural Network. In *Procs. of Knowledge Engineering System Conf.*, Crema, Italy, 2002.
- [26] T. Tsuji, H. Hattori, M. Watanabe, and N. Nagaoka. Development of Night-vision System. *IEEE Trans. on Intelligent Transportation Systems*, 3(3):203–209, Sept. 2002.
- [27] P. Viola, M. J. Jones, and D. Snow. Detecting Pedestrians using Patterns of Motion and Appearance. In *Procs. IEEE Intl. Conf. on Computer Vision*, pages 734–741, Nice, France, Sept. 2003.
- [28] C. Wöhler, J. K. Aulaf, T. Pörtner, and U. Franke. A Time Delay Neural Network Algorithm for Real-time Pedestrian Detection. In *Procs. IEEE Intelligent Vehicles Symposium '98*, pages 247–251, Stuttgart, Germany, Oct. 1998.
- [29] M. Yasuno, S. Ryouyuke, N. Yasuda, and M. Aoki. Pedestrian Detection and Tracking in Far Infrared Images. In *Procs. IEEE Intl. Conf. on Intelligent Transportation Systems 2005*, pages 13–16, Vienna, Austria, Sept. 2005.
- [30] L. Zhao. *Dressed Human Modeling, Detection, and Parts Localization*. Ph.D. dissertation, Carnegie Mellon University, 2001.
- [31] L. Zhao and C. Thorpe. Stereo and neural network-based pedestrian detection. *IEEE Trans. on Intelligent Transportation Systems*, 1(3):148–154, Sept. 2000.