Vision Algorithms and embedded solution for pedestrian detection with far infrared camera

Mircea Paul Muresan, Raluca Brehar, Sergiu Nedevschi
Computer Science Department
Technical University of Cluj Napoca, Romania
mircea.muresan, raluca.brehar, sergiu.nedevschi @ cs.utcluj.ro

***Abstract*— in the automotive industry the issue of safety remains a major priority. This aspect is not focused just on the driver but also on the other participants of the traffic like the pedestrians. This paper describes a pedestrian detection system where three different classification methods are used for detecting pedestrians with a far infrared camera. The three methods are tested and compared on variable number of features in order to obtain a scalable solution. The authors propose a low cost embedded implementation for the classification method that has proven to be best with respect to the accuracy and training time, taking the HOG as features descriptors for the region of interest**

Keywords— Pedestrian detection, Infrared Imagery, Driver Assistance, Artificial Neural Networks, ensemble algorithms, genetic algorithms, Histogram of oriented Gradients, embedded programming

I. Introduction

In the automotive industry the issue of safety of the driver and the other participants of the traffic remains a major concern. Most of the accidents are caused by human misjudgment or error [1] and it has been shown that more than two thirds of the accidents are happening during the night [2] or in bad visibility conditions. Pedestrian detection is currently a large research field and its applications can be used in the automotive industry but also in other areas like security and surveillance. The next generation driver assistance systems will help with the detection of pedestrians allowing the driver to focus on the road while making sure dangerous situations are avoided or mitigated. Due to the advancements in sensor technologies, the cost reduction of thermal sensors and the advantages they provide over the classical VL cameras the use of FLIR cameras has become a viable solution.

Some advantages of FLIR cameras that make them more suitable than the visible light cameras for solving the problem of detecting pedestrians especially during the night are:

- FIR cameras can detect in bad weather better than regular cameras including rainy and fogy weather
- Less susceptible to rays of other wave length
- Are susceptible to light such as head lights
- IR reduces the problem of shades
- In case of strong external heat radiation clothes that people wear can have different thermal behaviour depending on their type and color thus introducing textures to the image
- Do not require a light source
- Do not saturate when oncoming vehicles are present in the image
- Can reveal people, animals and obstructions on rural roads from long ranges. This early detection can improve reaction time
- The cost for thermal sensors has reduced dramatically

Some techniques of detecting pedestrians can be used from the area of detection with visible cameras. However mainly due to the differences between the two types of sensors many techniques are not applicable [3] and we have to come up with new methods for solving this issue.

II. Related work

There are very many approaches in detecting pedestrians with an IR camera. An interesting algorithm can be found at [4]. In this approach a three step process is taken in order to correctly find the pedestrians (contour based candidate extraction, tracking and classification). After the process of segmentation the pedestrian body parts that are considered to belong to the same pedestrian are grouped together to form the region of interest. The constrained condition of this algorithm is that the distance between each body part is the same. This technique uses two FIR cameras in order to find the distance to the objects of interest. Another interesting approach decomposes the image into two layers [5]. This layered approach has been widely used with VL cameras and it is interesting to see how such a method would behave with an infrared camera. The infrared image is decomposed in two parts the foreground (moving objects) and the background layer (still objects). The algorithm for p.d. is based on 2 things : shape and appearance. A SVM classifier are used to capture the variations in human shape for p.d. The features used in SVM are leanness and compactness of the ROI. Due to the fact that only the objects that were considered moving were assigned to the foreground the algorithm failed to detect still pedestrians. This problem made the authors to misclassify the pedestrians that stood still. An inspiring part based pedestrian detection technique was presented at the ITSC conference in 2013[6]. The authors have used two synchronized cameras for pedestrian detection, a VL camera and one FLIR camera. They have detected if a ROI is a pedestrian in one camera and that detection was validated with the help of the other camera. In this approach the authors split the sliding window into subparts and train a SVM for each part. The result of the SVM is then used in conjunction with the MLN. The final result of the detection is positive if at least one node in the MLN is true.

III. Detailed description of the pedestrian detection system

A. ROI Generation

It is very challenging to find a resource efficient method to generate our ROI since in the automotive domain there are many constraints regarding hardware size, power consumption etc. In order to eliminate information that we do not want we will first make a segmentation of our image. To segment our image we will make a dynamic threshold based on the histogram, taking the first value after the histogram mean.

Other methods for segmentation are possible. Another interesting approach that we have tried can be seen in [7] where the threshold value is chosen as the last local minimum before the saturation point. Our method however was more suitable for our benchmark and offered a better segmentation of the image. After we have segmented our image we will traverse it with a 64x128 pixel window. In case the ratio white pixels / black pixels in the ROI is > 0.30 the corresponding region is taken as a region of interest and passed to the next level in our pedestrian detection system.

B. Feature extraction

In our experiment we have tested our classifiers on variable number of features. First of all we have tested our classifiers on a small number of features (six features).

The six features used are the longest line between 2 white points in the segmented image and the slope of this line ,the mean and the standard deviation of the region of interest and the mean and the standard deviation of the image to which we have applied a Laplacian filter after smoothing it with a Gaussian filter. For the large number of features we have used histogram of the oriented gradients. The HOG descriptor has an advantage over other feature descriptors because it maintains its invariance to geometric and photometric transformations. This techniqe can be used for the detection of humans[8]. For the computation of the HOG descriptor we have used the HOGDescriptor from emgu cv. An implementation of the function that computes the descriptor can be found at [9].

C. ROI Classification

One of the most important aspect of a detection system is its robustness. The system needs to correctly classify the regions of interest because otherwise, in case it has too many false positives it is unreliable and can lead to the production of an accident while if it has too many false negatives it may be considered unreliable by the driver.

The classification methods used are: ADA Boost, artificial neural networks trained with back-propagation and artificial neural networks trained with genetic algorithms.

ADA Boost is an ensemble algorithm that creates a strong classifier from a weighted sum of week classifiers. We have done experiments with two different types of weak classifiers. The first classifier used was the decision stump and the second experiment was done with logistic regression. We have taken the number of weak learner equal to the number features when we have used the HOG descriptor and for the 6 features we have used 45 weak classifiers.The other experiment we have done with HOG features uses logistic regression as a weak classifier. We use the logistic function (1)

$$f(x) = \frac{1}{1+e^{-x}} \quad (1)$$

2. Artificial neural networks trained with back-propagation

The neural network topology used is that of a multilayer feed forward neural network. The network is first initialized with random weights, the gradient of the error function is chosen and it is used to correct the initial weights. Back propagation is the name of the algorithm used to correct the initial weights. This training procedure is based on the gradient descent model which looks for the minimum of the error function. The solution to the problem is the neural network whose weights that minimize the error. We have used the sigmoid function from (1) as a squashing function.

As our network architecture we have used 3 layer artificial neural net. For the experiments that we have done on small amount of features we have used 6 unit input layer, 20 unit hidden and 1 unit output. We have tried various architectures however this type of architecture gave us the best results. We have chosen as a learning rate the value of $\eta = 0.001$ and a momentum of 0.8.

For the case in which we have used the histogram of oriented gradients the input layer had a dimensionality of 3780 activation units, the hidden layer 3781 and the output layer had 1 unit. For this network we have used a learning rate of 0.6 and a momentum of 0.9. More details on neural networks and also how to implement them can be found at [10, 11].

3. Artificial neural networks generated with genetic algorithms

Another method with which we have trained our neural classifier is by using genetic algorithms. These algorithms are based on Darwin’s theory of evolution through natural selection. The organisms that are fit have the privilege to mate and produce more adaptable offspring. Each individual can be taught as a solution to the problem we are trying to solve. In our experiment the individual is a feed forward neural network. The network has been flattened in order to better make the genetic operations on it. Each gene is a synaptic weight of a perceptron. Once the genetic operations are done we convert the network back to its original format.

The process in which new individuals are created is a three step process. First the suitability of the individual to the environment is tested through a fitness function and the individuals that are considered to be the fittest are selected to mate. The fitness function in our case is composed by the difference between the actual – ideal outputs of a network to our input set, the smaller the error the best is the individual. There are several types of selection mechanisms [12], in our experiment however, we used ranking selection. The mating of the individuals is the second step and this is done through the process of crossover. In our solution we have used a double point crossover to mate two individuals [13]. Finally in order to avoid having duplicates and to avoid getting stuck in a local minima solution we use the process of mutation. In our solution we take a gene from our chromosome and multiply it with a ratio so it changes randomly (2) and (3).

$$ratio = random(range) - range \quad (2)$$

$$gene = gene * ratio \quad (3)$$

The range specifies how large should the mutation be allowed. In order to have a good performance we have kept the architecture of the neural network from the previous point and we use a population of 4000 individuals to cover our search space. The mutation percent was of 0.1 and the crossover percent was of 0.3.

IV. EXPERIMENTAL RESULTS

In this section we will report our experimental results using the images from the git repository[14]. The benchmark covers a variaty of conditions such as rainy, clowdy, foggy periouds and the pedestrians are occluded, overlaped and also not occluded. Our trainning set contained 1536 positive images and 3910 negative images. The test set contained 20000 images out of which aproximatively half were with pedestrians and half were with non pedestrians. One important thing to note is that our non pedestrian far infrared data base does not contain images with animals. We will analyze the pedestrian detection algorithms first with the six extracted features and then using the HOG descriptor. In order to evaluate our algorithms we use the precision,recall and accuracy metric. All the algorithms have been trained on the same data and also the test set was the same for the three algorithms so that we can better compare them. The algorithms were run on an intel i5 computer that had 2GB DDR3.

A. Analysis of the algorithms using the six features.

For our multilayer neural network architecture we have found a learning rate experimentally which is of 0.001. The number of epochs for the algorithm was 4000. The time needed for the training was of aproximately 3 hours.

The values for the detection are : true positives – 87.154% , false positives – 12,845%, true negatives – 10,92%, false negatives – 89,07%. The precision, recall and accuracy values are : 0.871, 0.133, 0.212.

The adaboost classifier was designed first using decision stump as weak classifier and then for better accuracy logistic regression. The number of iterations was 45 and in each iteration a new weak classifier was learned and added to the final decision rule. The execution time of the training was under 5 minutes. This algorithm had a true positive percentage of 93,008% and false positives of 6.991%. The values for true negatives are 61.023% and false negatives 38.976%. The precision, recall and accuracy values are: 0.930, 0.272, 0.653. For the neural networks generated with genetic algorithms we have decided to use the same network topology as in the artificial neural network trained with backpropagation.We have used a population of 4000 individuals and 400 epochs. Even though we have paralelized the algorithm so it could run much faster it has converged in aproximatively 2 days. The results are : tp – 68,94%, fp – 31.05%, tn – 65.70%, fn – 34.3%. The precision, recall and accuracy values for the genetic algorithm ar : 0.68, 0.184 and 0.54. In figure 4 and figure 5 we have a snapshot where we can see the classification results of the adaboost on an infrared image with pedestrians.

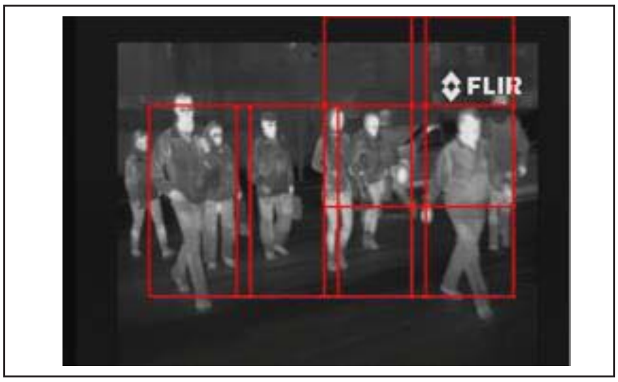

Fig. 4. Snapshot of the program detecting all the pedestrians from an IR image

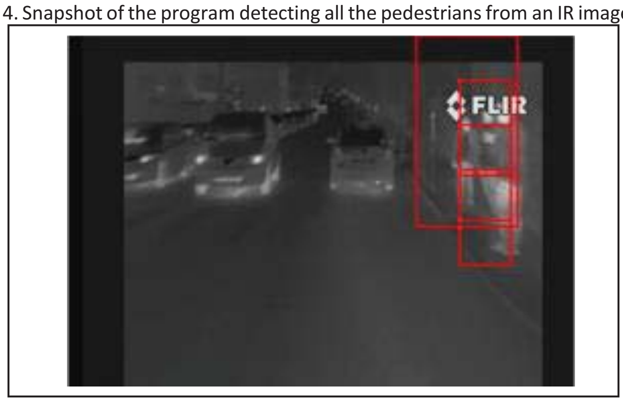

Fig. 5. Snapshot of the program detecting pedestrians using ADA Boost with decision stump classifier

B. Analysis of the algorithm using HOG

In this part we will analyze the performance of the classifiers using large number of features.

For the first classifier, the artifical neural network, we have changed the topology(3780, 3781 and 1) and learning rate(0.1) to better fit the large amounts of features. The network took 4 hours to train and the results are the following tp – 94,47%, fp – 5.53% , tn – 10.15%, fn – 89.84%. The precision, recall and accuracy values for this algorithm : 9.44, 0.14, 0.23.

For the ADA Boost classifier we have choosen a number of weak learners equal with the number of features obtained from HOG(3780). The classifier was trained in approximatively one hour and the results are given can be seen bellow.

The algorithm had a tp rate of 92.92%, and a fp percentage of 7.07%. The tn value is – 44.444% and fn – 55.555%. The values for precision, recall and accuracy are : 0.929, 0.45 and 0.59. The evaluation values for the ADA Boost algorithm with HOG features are true positive - 97.5%, false positives – 2.5 %, true negatives – 28% and false negatives- 72%. The precision, recall and accuracy value of the ADA Boost algorithm using logistic regression are : 0.97, 0.57, 0.627.

As we can see from the results, even though the artificial neural networks have a good precision the overall accuracy of the algorithm is smaller compared to ADA Boost mainly because the algorithm has many false negatives.

The neural network generated with genetic algorithms was the least accurate. Since we had a very high dimensionality of the features we could create just 6 individuals to cover an extremely large search space. We could not increase the number of individuals without having memory leaks during the training process. For the high dimensionality of the features this classifier has behaved very bad. The results are tp – 26%, fp – 73.98%, tn – 5.48%, fn – 94.52%. The precision, recall and accuracy values : 0.26, 0.41, 0.095.

The experimental results show us that the ensemble algorithm, ADA Boost, outperforms the other two classifiers on large amount of features with respect to the training time and accuracy. We can also remark that the neural networks generated with genetic algorithms are good classifiers if the search space is not very big outperforming in accuracy the neural networks. Out of the two versions of ADA Boost, the classifier trained with logistic regression performs the best in terms of accuracy.

V. EMBEDED BOARD CHARACTERISTICS

The olinuxino a13 board is a low cost linux computer development board. This board is an open source hardware and software board which is noise imune and whose GPIOs operate in industrial environment (-25 +85 C). There is no restriction in manufacturing, selling or reselling these boards. The board uses a Cortex A8 microcontroller which is very competitive for Android tablets giving higher performance at low cost. The processor makes multitasking smoother and responsive touch better. Detailed characteristics of the board and its processor can be found at [15, 16]. The board allows support for several computer vision libraries like open cv and scikit-learn and also all the major programming languages like python, c++ etc.

We have implemented ADA Boost with logistic regression on the board since it had the best results (memory consumption and accuracy). We have also attached a visual and acoustic warning system to the board in order to better warn the driver. The results of the detection are also shown on a lcd module that was attached to the board.

VI. CONCLUSIONS AND FURTHER STUDY

In this paper we have implemented a pedestrian detection method and we have tested and compared three methods for classifying pedestrians using an infrared camera. The three methods were compared on variable number of features. The classifier that has proven to be the best with respect to the HOG features, in terms of accuracy, and memory consumption was ADA Boost. Training an ANN classifier with genetic algorithms has proven to be better than the training with back propagation, even though the second had a larger number of true positives. ADA Boost with logistic regression was implemented on a low cost embedded board giving good results in terms of classification accuracy. This shows that a pedestrian detection systems can be implemented on small embedded board and the costs of such a system would not be very high.

The authors will try to create a stronger classifier by combining the genetic algorithm with ADA Boost since the genetic algorithm on small number of features proved to give a good performance. The authors will also try to develop a part based pedestrian detection method using human body proportions and an ADA Boost classifier will be used to train each individual part. The non-pedestrian data base will be improved with images of animals. Future algorithms will also try to take into account the animals that may appear on the road.

ACKNOWLEDGEMENTS

This work has been partially supported by UEFISCDI (Romanian National Research Agency) in the CELTIC+ research project Co-operative Mobility Services of the Future (CoMoSeF).

REFERENCES

[1] P. Sa;mon, M. Regan and I Johnston. Human error and road transport. Monash University, 2005. Available 31st March 2008 http://www.monash.edu.au/muarc257.pdf

[2] *Pedestrian Safety*. Irish Road Safety Authority,2005. Available 31st March 2008: http://www.rsa.ie/publication/publication/upload/pedestrian safety.doc.

[3] Y. Fang, K. Yamada, Y. Ninomiya, B. Horn, and I. Masaki. Comparison between infrared-image-based and visible-image-based approaches for pedestrian detection. In *Proceedings* IEEE Intelligent Vehicles Symposium,pages 505–510, 2003.

[4] Hiroaki SAITO, Takeshi HAGIHARA, Kenichi HATANAKA and Takanori SAWAI Development of Pedestrian Detection System Using Far- Infrared Ray Camera," SEI Technical Reviw Vol.171,80-85, 2007

[5] Congxia Dai, Yunfei Zheng and Xin Li Lane Department of Computer Science and Electrical Engineering West Virginia University Morgantown, WV 26506-6109 congxiad,zheng,xinl@csee.wvu.edu, 2005

[6] Daniel Olmeda, Jose Maria Armingol and Arturo de la Escalera , Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems (ITSC 2013), The Hague, The Netherlands, Part based Pedestrian Detection Based on Logic Inference October 6-9, 2013

[7] U. Meis, W. Ritter, and H. Neumann. Detection and classification of obstacles in night vision traffic scenes based on infrared imagery. In Proceedings IEEE Intelligent Transportation *Systems*, volume 2, pages 1140–1144 vol.2, 2003.

[8] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. In CVPR, pages 886-893, 2005

[9] http://stackoverflow.com/questions/20497798/taking-the-hog-descriptor-of-an-image-using-hogdescriptor-from-emgu-cv-c-sharp

[10] R. Rojas: Neural Networks, Springer-Verlag, Berlin,1996

[11] Jeff Heaton Introduction to Neural Networks for C#, 2nd Edition

[12] Research Scholar, Assistant Professor (Senior Grade), Department Of Computer Science and Engineering, Velalar College of Engineering and Technology, Erode, Tamil Nadu, India. Dr.T.RAVICHANDRAN Principal and Research Supervisor, A REVIEW OF SELECTION METHODS IN GENETIC ALGORITHMR.SIVARAJ, Hindusthan Institute of Technology Coimbatore-641032, Tamil Nadu, India, 2012

[13] JORGE MAGALHÃES MENDES Department of Civil Engineering CIDEM School of Engineering –Polytechnic of Porto Rua Dr. António Bernardino de Almeida, 431 –4200-072 Porto PORTUGAL jjm@isep.ipp.pt A Comparative Study of Crossover Operators for Genetic Algorithms to Solve the Job Shop Scheduling Problem, 1995

[14] https://github.com/MMp131316/IR.git

[15] http://www.farnell.com/datasheets/1696781.pdf - datasheet

[16] https://www.olimex.com/Products/OLinuXino/A13/A13-OLinuXino/resources/A13-Brief.pdf