Improvised Pedestrian Detection Algorithm

Zainab Mudassir Abdul Azeez
Department of Computer Science
Sri Jayachamarajendra College of Engineering
Karnataka, Mysore 570004
Email:zainabmudassir@gmail.com

Thejasweni Prakash M
Department of Computer Science
Sri Jayachamarajendra College of Engineering
Karnataka, Mysore 570004
Email:thejasweniprakash@gmail.com

***Abstract*—Due to the progressive increase in population, the daily means of transportation requires ways to control and interpret the obstacles encountered. A technology which incorporates deep learning models can be used in smart vehicles. This pedestrian detection algorithm performs in near real- time and with higher accuracy than existing systems. A typical detection algorithm breaks down an image into many windows which must be inspected by video frame at speeds ranging from 5-30 frames per second. In cascade detection, the detector operates through three stages, each stage looking at a window with different perspective. Traditional cascade detection relies on weak learners, which are simple classifiers in all the stages. But, these classifiers are incapable of performing highly complex classification. The solution to this is an algorithm that implements deep learning models in the final stages of a cascaded detector which makes it capable of optimizing the trade-off between detection accuracy and speed for cascades with stages of such different complexities. These are the first cascades to include stages of deep learning. The results obtained with this new algorithm are substantially better for real-time, accurate pedestrian detection. This paper also includes the combination of a collision avoidance system along with the above stated algorithm.**

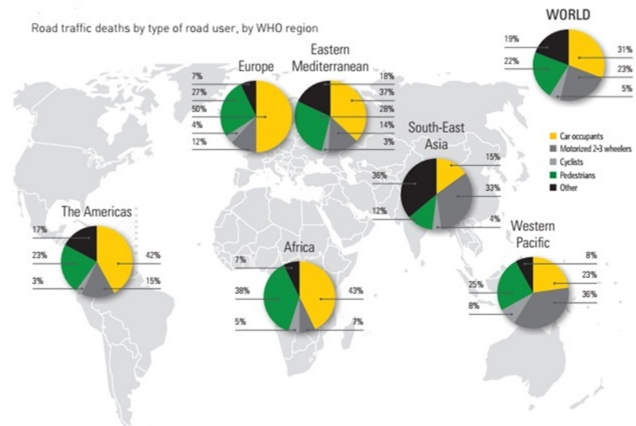

Fig. 1.

I. Introduction

Pedestrian accidents are one of the major sources of traffic injuries and fatalities. Fig.1. shows us how much pedestrian safety is necessary since the death rate and threat to lives is considerably high number due to the vehicle accidents. Therefore, in order to avoid collisions with pedestrians they must be detected, being camera sensors key due to the rich amount of cues and high resolution they provide. This paper brings together a new algorithm and insight to construct a framework for a robust pedestrian detection. A typical pedestrian detection algorithm breaks down an image into small windows that are processed by a classifier that signals the presence or absence of a pedestrian. This approach is challenging because pedestrians appear in different sizes–depending on distance to the camera– and locations within an image.

In cascade detection, the detector operates throughout a series of stages. In each of these stages, the cascade detectors employ weak learning models. The complexity increases as the algorithm clears levels and so does the number of weak learners applied. In the first stages, the algorithm quickly identifies and discards windows that it can easily recognize as not containing a person (such as the sky) implying the least complexity amongst all the levels. The next stages process the windows that are harder for the algorithm to classify, such as those containing a tree, which the algorithm could recognize as having person-like features (shape, color, contours, etc.). In the final stages, the algorithm must distinguish between a pedestrian and very similar objects. However, because the final stages only process a few windows, the overall complexity is low[1]. While this method is fast, it isn't powerful enough when it reaches the final stages because the weak learners used in all stages of the cascade are identical. Therefore the last stages are not necessarily capable of performing highly complex classification. A new novel algorithm incorporates deep learning models in the final stages of a cascaded detector. They are better suited for complex pattern recognition, which they can perform after being trained with hundreds or thousands of examples. However, they work well for the final cascade stages, they are too complex to be used in the early ones. The solution is a new cascade architecture that combines classifiers from different families: simple classifiers (weak learners) in the early stages complex classifiers (deep learning models) in the later stages. No previous algorithms have been capable of optimizing the trade-off between detection accuracy and speed for cascades with stages of such different complexities. In fact, these are the first cascades to include stages of deep learning.

The results obtained with this new algorithm are substantially better for real-time, accurate pedestrian detection[2]. After achieving great speed and accuracy, the algorithm can be further improved by combining it with an independent system of a type of collision avoidance. In this paper we have proposed a collision avoidance system which not only takes care of the possible collision with the pedestrian but also with the other aspects in the frame. For example, rear-end collisions caused by unannouned deceleration or braking of the preceding vehicle.

II. Related work

Pedestrian detection has been a central topic in computer vision research, spanning morethan 20 years of research [10,15,17,19]. A wide variety of methods have been applied topedestrian detection over the years, with continued improvement in performance [10,12,18,20,23]. Some methods focus on improving the base features used, whereas others focus on the learning algorithms [13,14], or other techniques such as incorporating Deformable Parts Models [18,21] or using context [13,16,21,22]. Dollaret al. [12] developed a benchmark and evaluation toolbox that has been instrumental intracking progress in the field. Benenson et al. [10] have recently proposed a comparative paper that evaluates performance of various features and methods on pedestrian detection. Viola and Jones proposed a cascade-of-classifiers approach [20], which has been widely used for real-time applications. The method has been extended by employing different types of features and techniques [13], but fundamentally the concept of the cascade, with early rejection of majority of test examples, has been widely utilized to achieve real-time performance. Perhaps the most popular feature used for pedestrian detection (and several other image-based detection tasks) is the HOG feature developed by Dalal and Triggs [11]. Although not real-time, about 1 FPS, this work has been instrumental to the development of faster and more accurate features for pedestrian detection, which are used in the top performing methods in combination with SVM or Decision forests [9,12].

Deformable Parts Models [17] have shown success on the pedestrian detection task [18,21]. Deep learning-based techniques have also been applied to pedestrian detection and have led to improvements in accuracy [14,16]. These approaches are still slow, ranging from over a second per image [14] to several minutes [20]. The faster approaches do not apply deep nets to the raw pixel input so theri accuracy is reduced. Improving the speed of pedestrian detection has also been an active area. Benson et al proposed a method reaching speeds of 100 to 135 FPS for detection in a 480x640 image, albeit with significantly lower accuracy. Other researchers have focused specifically on speeding up Deep Neural Networks, but with no real-time solutions yet.

III. Multistage Detection

A. Cascaded Detectors

The cascade classifier consists of stages, where each stage is an ensemble of weak learners. The weak learners are simple

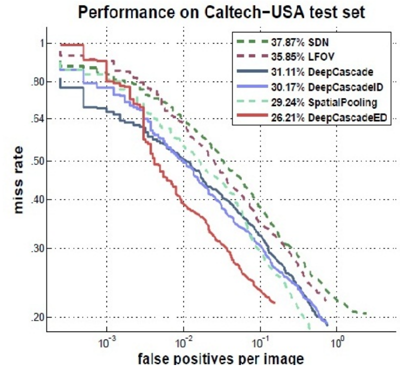

Fig. 2. Results of our DeepCascade methods on the Caltech test data for pedestrians. The best state-of-the-art method (SpatialPooling) and the best state-of-the-art Deep Networks methods SDN, LFOV are shown for comparison. All other methods are shown in Table 1. Our network performs much better in the area of the curve with small number of false positives, which is the target area for a practical pedestrian detection system.[36]

classifiers called decision stumps. Each stage is trained using a technique called boosting. Boosting provides the ability to train a highly accurate classifier by taking a weighted average of the decisions made by the weak learners.

Each stage of the classifier labels the region defined by the current location of the sliding window as either positive or negative. Positive indicates that an object was found and negative indicates no objects were found. If the label is negative, the classification of this region is complete, and the detector slides the window to the next location. If the label is positive, the classifier passes the region to the next stage. The detector reports an object found at the current window location when the final stage classifies the region as positive.

The stages are designed to reject negative samples as fast as possible. The assumption is that the vast majority of windows do not contain the object of interest. Conversely, true positives are rare and worth taking the time to verify.

A true positive occurs when a positive sample is correctly classified. A false positive occurs when a negative sample is mistakenly classified as positive. A false negative occurs when a positive sample is mistakenly classified as negative. To work well, each stage in the cascade must have a low false negative rate. If a stage incorrectly labels an object as negative, the classification stops, and you cannot correct the mistake. However, each stage can have a high false positive rate. Even if the detector incorrectly labels a nonobject as positive, you can correct the mistake in subsequent stages.

The overall false positive rate of the cascade classifier is fs, where f is the false positive rate per stage in the range (0 1), and s is the number of stages. Similarly, the overall true positive rate is ts, where t is the true positive rate per stage in the range (0 1]. Thus, adding more stages reduces the overall false positive rate, but it also reduces the overall true positive rate[24]

B. Deep Learning Models

Deep learning (deep structured learning, hierarchical learning or deep machine learning) is a branch of machine learning-based on a set of algorithms that attempt to model high-level abstractions in data by using multiple processing layers, with complex structures or otherwise, composed of multiple non-linear transformations.[25][26][27][28][29][30]

Deep learning is part of a broader family of machine learning methods based on learning representations of data. An observation (e.g., an image) can be represented in many ways such as a vector of intensity values per pixel, or in a more abstract way as a set of edges, regions of particular shape, etc. Some representations are better than others at simplifying the learning task (e.g., face recognition or facial expression recognition[31]) from examples. One of the promises of deep learning is replacing handcraftedfeatures with efficient algorithms for unsupervised or semi-supervised feature learning and hierarchical feature extraction.[32]

There are a number of ways that the field of deep learning has been characterized. Deep learning is a class of machine learning algorithms that[25](pp199200)

use a cascade of many layers of nonlinear processing units for feature extraction and transformation. Each successive layer uses the output from the previous layer as input. The algorithms may be supervised or unsupervised and applications include pattern analysis (unsupervised) and classification (supervised). are based on the (unsupervised) learning of multiple levels of features or representations of the data. Higher level features are derived from lower level features to form a hierarchical representation. are part of the broader machine learning field of learning representations of data. learn multiple levels of representations that correspond to different levels of abstraction; the levels form a hierarchy of concepts.

Deep neural networks[edit] A deep neural network (DNN) is an artificial neural network (ANN) with multiple hidden layers of units between the input and output layers.[26][28] Similar to shallow ANNs, DNNs can model complex non-linear relationships. DNN architectures, e.g., for object detection and parsing generate compositional models where the object is expressed as a layered composition of image primitives.[33] The extra layers enable composition of features from lower layers, giving the potential of modeling complex data with fewer units than a similarly performing shallow network.[26]

DNNs are typically designed as feedforward networks, but recent research has successfully applied the deep learning architecture to recurrent neural networks for applications such as language modeling.[34] Convolutional deep neural networks (CNNs) are used in computer vision where their success is well-documented.[35]

IV. Proposed Pre Collision System

A collision avoidance system is an automobile safety system designed to reduce the severity of an accident. Also known as precrash system, forward collision warning system or collision mitigating system, it uses radar and sometimes laser and camera sensors to detect an imminent crash. The presence of a pre-collision system in vehicles is not a new concept as of today. Although pedestrian detection has contributed in avoiding and controlling accidents with pedestrians only n the case where the driver is alert and keeps a constant check on the visual update of the incoming traffic (in this case the pedestrians) given by the detection algorithm. But, in situations where the driver is not alert, sleepy or inattentive the detection algorithm fails to contribute enough since the driver does not notice the detected pedestrian he is likely to encounter. To overcome this drawback, a pedestrian detection algorithm can employ a collision avoidance system along with it, which is basically a combination of two independent systems. Recently, Toyota implemented this concept and it works as follows. If the pedestrian is at a distance too close to the incoming vehicle, the vehicle initiates automatic brakes regardless of attentiveness of the driver. This technology although ensures safety of the pedestrians, puts at risk the security of the vehicles behind it. In a situation where the vehicles are allowed high speed limits and an encounter with jaywalkers is a possibility, such automatic breaks can turn out to be a threat for the vehicles behind since halting of the car is a surprise to them like in Fig. 3. Here, the car with velocity Vo will show no signs of slowing down as it has no idea that the car with velocity Vp is going to decelerate/halt result of which will be covering of the distance D and hence the collision. This might lead in a collision as a result of avoiding the earlier collision. To overcome this, we came up with a pre collision system, that instead of applying sudden automatic breaks, after the detection of the pedestrian, calculates the distance between them and decides if it can be avoided by just slowing down the velocity of the vehicle. It uses the concept of automatic brakes only when the situation is inevitable.

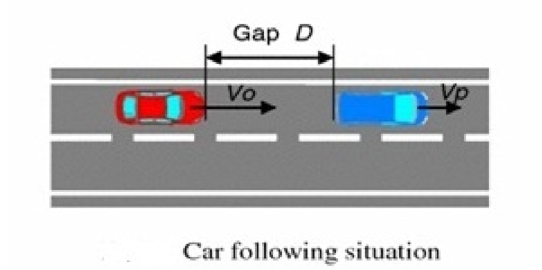

Fig. 3.

Recently, several research studies have been conducted on driver's deceleration behaviours including collision risk for application to driver assistance system. Kondoh et al. Investigate the risk perception and showed that it can be represented by TTC (Time-to-collision) and THW (Time-headway) [6]. Isaji et al. [7] and Wada et al. [8] have proposed a performance index of approach and alienation, as a model of driver's perceptual risk of a preceding vehicle. A deceleration assistance control method will be proposed for preventing rear-end crash based on expert driver's deceleration model derived from driver's perceptual risk. Initiation timing of brake assistance will be determined by driver's brake initiation timing model. [5]

Generally, a collision avoidance system transmits an intermittent radar beam (so it sends a signal only part of the time) and, for the rest of the time, "listens" out for any reflections of that beam from nearby objects. If reflections are detected, it knows something is nearbyand it can use the time taken for the reflections to arrive to figure out how far away it is[3]. But, in the system proposed, it sends the radio waves towards the pedestrian to calculate the distance between them. Advanced Driver Assistance Systems (ADAS) refer to the set of smart components such as object detection and tracking systems used to protect drivers and road users. An important task that should be performed by ADAS is the measurement of the distance between a vehicle and a detected object. This helps to anticipate the position of objects in the lane, prevents collisions and improves the safety on the road. For this purpose, Distance Measurement Systems (DMS) including Vision-based techniques, Millimeter wave radars, Infrared ranging, are used in the automotive industry[4]. The determined distance in turn can calculate the time remaining for collision with the pedestrian once the vehicle has crossed the threshold distance of safety. The system will make an attempt to warn the driver both visually and acoustically. If the driver is still not undertaking measures, the system invokes the vehicle to slow down at a particular rate automatically, depending on the time and distance remaining for the accident to occur.

V. Conclusion

The above proposed approach for pedestrian detection aims to not only increase the speed of evaluating the frame of the image but to do so by not putting the accuracy at risk. To achieve this, a Deep Learning Model-based algorithm for pedestrian detection which combines the ideas of fast cascade and a deep network has been presented here. It is simple to implement as it is based on open source implementations. This method ranks among the best ones for pedestrian detection and runs in real-time, at 15 frames per second. This is the only method we are aware of that is both real-time and achieves high accuracy. To increase the quality of the algorithm it has been combined with a collision avoidance system which takes care of both the pedestrian and also the vehicles behind it by avoiding rear-end collisions through a deceleration concept. We hope this method will help future works to continue to improve pedestrian detectors in terms of both accuracy and speed and also reaches amazing heights with respect to safety so that more methods can be usable for pedestrian detection for real-time applications. Future work can include increasing the depth of the deep networks cascade by adding more tiny deep networks and exploring the efficiency-accuracy trade-offs. The essence of the pre collision system adds importance to the algorithm as it is making it more useful. This approach aims to bring down the number of fatalities and casualties caused by pedestrian accidents and also prevents a possible rear-end collision which could further worsen the situation.

References

[1] https://www.sciencedaily.com/releases/2016/02/160208095903.htm
[2] http://jacobsschool.ucsd.edu/news/news_releases/release.sfe?id=1883
[3] http://auto.howstuffworks.com/car-driving-safety/safety-regulatory-devices/pre-collision-systems.htm
[4] http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=arnumber=7266486url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D7266486
[5] http://ieeexplore.ieee.org/xpl/login.jsp?tp=arnumber=5354640url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D5354640
[6] Kondoh, T., Yamamura, T., Kitazaki, S., Kuge, N., Boer, E.R.: Identification of Visual Cues and Quantification of Drivers Perception of Proximity Risk to the Lead Vehicle in Car-Following Situations.In: Proc. of Driving Simulation Conference Asia/Pacific (2006)
[7] Isaji, K., Tsuru, N., Wada, T., Imai, K., Doi, S., Kaneko, H.: Emergency Evaluation by Approaching in Longitudinal Direction (First Report) Proposal of Performance Index for Approach and Alienation based on the Driver Operations. Transactions of Society of Automotive Engineers of Japan 38(2), 2530 (2007) (in Japanese)
[8] Wada, T., Doi, S., Imai, K., Tsuru, N., Isaji, K., Kaneko, H.: Analysis of Drivers Behaviors in Car Following Based on A Performance Index for Approach and Alienation. SAE 2007 Transactions, Journal of Passenger Cars Mechanical Systems, SAE paper 2007-01-0440, 306320 (2008)
[9] R. Benenson, M. Matthias, T. Tuytrlaars, and L. Van Gool. Seeking the strongest rigid detector. CVPR, 2013.
[10] R. Benenson, M. Omran, J. Hosang, and B. Schiele. Ten years of pedestrian detection,what have we learned? 2nd Workshop on Road scene understanding and Autonomous driving, ECCV, 2014.
[11] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. CVPR, 2005.
[12] P. Dollar, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: A benchmark.CVPR, 2009.
[13] P. Dollar, R. Appel, and P Perona. Crosstalk cascades for frame-rate pedestrian detector. ECCV, 2010.
[14] P. Luo, X. Zeng, X. Wang, and X. Tang. Switchable deep network for pedestrian detection. CVPR, 2014.
[15] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio. Pedestrian detection using wavelet templates.CVPR, 1997.
[16] W. Ouyang and X. Wang. Joint deep learning for pedestrian detection. ICCV, 2013., 2014.
[17] C. Papageorgiou, T. Evgeniou, and T. Poggio. A trainable pedestrian detection system. Proceedings of Intelligent Vehicles, Stuttgart, Germany, 1998.
[18] D. Park, D. Ramanan, and C. Folwkes. Multiresolution models for object detection. ECCV, 2010.
[19] K. Rohr. Incremental recognition of pedestrians from image sequences. CVPR, 1993.
[20] P. Viola, M. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance.ICCV, 2003.
[21] J. Yan, X. Zhang, Z. Lei, S.Liao, and S. Li. Robust multi-resolution pedestrian detectionin trafn Ac scenes.CVPR, 2013.
[22] X. Zeng, W. Ouyang, and X. Wang. Multi-stage contextual deep learning for pedestriandetection. ICCV, 2013.
[23] S. Zhang, C. Bauckhage, and A. Cremers. Informed haar-like features improve pedes-trian detection. CVPR, 2014.
[24] http://in.mathworks.com/help/vision/ug/train-a-cascade-object-detector.html?requestedDomain=www.mathworks.com
[25] Deng, L.; Yu, D. (2014). "Deep Learning: Methods and Applications"(PDF). Foundations and Trends in Signal Processing 7: 34.
[26] Bengio, Yoshua (2009). "Learning Deep Architectures for AI" (PDF).Foundations and Trends in Machine Learning 2 (1): 1127.
[27] Bengio, Y.; Courville, A.; Vincent, P. (2013). "Representation Learning: A Review and New Perspectives". IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (8): 17981828.
[28] Schmidhuber, J. (2015). "Deep Learning in Neural Networks: An Overview".Neural Networks 61: 85117.
[29] Bengio, Yoshua; LeCun, Yann; Hinton, Geoffrey (2015). "Deep Learning". Nature 521: 436444.

[30] Deep Machine Learning A New Frontier in Artificial Intelligence Research a survey paper by Itamar Arel, Derek C. Rose, and Thomas P. Karnowski. IEEE Computational Intelligence Magazine, 2013

[31] Glauner, P. (2015). Deep Convolutional Neural Networks for Smile Recognition (MSc Thesis). Imperial College London, Department of Computing.

[32] Song, H.A.; Lee, S. Y. (2013). ”Hierarchical Representation Using NMF”. Neural Information Processing. Lectures Notes in Computer Sciences 8226. Springer Berlin Heidelberg. pp. 466473.

[33] Szegedy, Christian, Alexander Toshev, and Dumitru Erhan. ”Deep neural networks for object detection.” Advances in Neural Information Processing Systems. 2013.

[34] T. Mikolov et al., ”Recurrent neural network based language model,” Interspeech, 2010.

[35] LeCun, Y.; et al. ”Gradient-based learning applied to document recognition”. Proceedings of the IEEE 86 (11): 22782324.

[36] Angelova ;et al.: Real-time pedestrian detection with deep cascades.