Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems (ITSC 2013), The Hague, The Netherlands, October 6-9, 2013

Partially Occluded Pedestrian Classification using Part-based Classifiers and Restricted Boltzmann Machine Model

Saleh Aly[1], Loay Hassan[1], Alaa Sagheer[2], and Hiroshi Murase[3]

Abstract— **One of the main challenges in pedestrian detection is occlusion. This paper presents a new method for pedestrian classification with partial occlusion handling. The proposed system involves a set of component-based classifiers trained on features derived from non-occluded dataset. The scores of all component classifiers are statistically modeled to estimate the final score of pedestrian. A generative stochastic neural network model namely Restricted Boltzmann Machine (RBM) is learned to estimate the posterior probability of pedestrian given its components scores. The training data used to train RBM model is artificially generated occluded data which simulate real occlusion conditions appeared in pedestrians. Experimental results on real-world dataset, with both partially occluded and non-occluded data shows the effectiveness of the proposed method.**

I. INTRODUCTION

Pedestrian detection based on computer vision algorithms acts an important role in many applications such as driver assistance system and surveillance system [1], [2], [3], [4]. Since pedestrian detection is one of the important components in driver assistance system, it has attracted much attention in recent years.

Pedestrian detection is more complicated than any other object detection problems like cars and traffic signs because of the high articulation of human body. The appearance of human depends on viewpoint, illumination, clothing, and occlusion. We propose a new method to overcome the partial occlusion problem using part-based representation of human. This representation helps to overcome some of the occlusion and changes in perceived 2-D pedestrian images.

Many classification approaches, features and deformation models have been used for achieving good results in object detection. Recently, Dollar et al. [5] evaluate 16 types of state-of-the-art pedestrian detection methods. Nearly all modern detectors employ some forms of histogram of oriented gradients (HOG) feature. In addition, detectors can utilize gradients directly, Haar-like, color, texture, self-similarity, integral histogram, local binary patterns, and motion feature. Most of the classifiers of these methods adopt Adaboost and linear support vector machine (SVM) [6] and a few adopt latent SVM [7] and Hik-SVM [8]. The experiments include 6 types of pedestrian databases, which is divided into near, middle, and remote groups for experiment. The best overall performing detector is integral channel features (CHNFTRS) [9] and the fastest pedestrian detector in the west (FDW) [10] and both methods use HOG, gradient, Haar-like features and Adaboost classifier. The worst one is the method using Haar-like features and Adaboost classifier [11].

Sliding window classifiers are presently the predominant method being used in pedestrian detection, due to their good performance. For sliding window detection approach, each image is densely scanned from the top left to the bottom right with rectangular sliding windows in different scales. For each sliding window, certain features such as edges, image patches, and wavelet coefficients are extracted and fed to a classifier, which is trained offline using labelled training data. The classifier will classify the sliding windows which contains person as positive samples and the other as negative samples. Currently, Support Vector Machine and variants of boosting decision trees are two dominant classifiers for their efficiency.

One of the major drawback of sliding window approach is its poor handling of partial occlusion due to its dense selection. If a portion of the scanning window is occluded, the features corresponding to the occluded area will be inherently noisy and will deteriorate the classification result of the whole window. On the other side, part-based detectors can alleviate the occlusion problem to some extent by relying on the non-occluded part to determine the human position. However, there is a key problem to be solved how to integrate the scores of part detectors when there are occluded. To handle the weakness of part detector and to combine the global information of full-body detector concurrently. There are various methods used to combine detection scores. In [12], the scores are thresholded and combined, however in [3], they use linear SVM to predict the final score when only intensity information is available. In [13], other cues like depth and motion are used as a complementary information for of intensity information.

Most of the state-of-the-art methods which handle occlusion start by estimating visible parts of the pedestrian and use these parts to estimate the final score [13], [12]. In this paper, a set of semantically and overlapped part-based SVM models are used to obtain the scores of component classifiers. Unlike previous occlusion handling approaches that depends only on the training of non-occluded data. A stochastic neural network model is trained to estimate the final score using a set of artificially generated occluded pedestrians. The occluded patterns generated by assuming that lower-body parts of pedestrian are highly probable to

[1] S. Aly is with Faculty of Engineering, Department of Electrical Engineering, Aswan University, Egypt. (email: saleh@aswu.edu.eg, loay@cairo.aswu.edu.eg)

[2] A. Sagheer is with Faculty of Science, Department of Mathematics, Aswan University, Egypt. (email:alaa@cairo.aswu.edu.eg)

[3] H. Murase is with Graduate School of Information Science, Department of Media Science, Nagoya University, Nagoya, Japan (email: murase@is.nagoya-u.ac.jp)

978-1-4799-2914-613/$31.00 ©2013 IEEE

be occluded than upper-body parts. The occlusion patterns used to generate occluded samples are derived from the contextual pixels around pedestrians in the non-occluded training images. The main advantages of this approach is that the final decision is based on a stochastic model learned from artificially occluded data

This paper is organized as follows, section II describes the works related to the proposed method, and section III explains the details of the proposed occlusion invariant pedestrian detection system. Section IV shows the experimental setup and results using both real occluded and non-occluded images. Finally, conclusions are described in section V.

II. Related Work

Wu and Nevatia [14] proposed a method for handling partial occlusion in human detection by modelling human as an assembly of natural body parts. They introduced edgelet features to describe silhouette oriented features. The response of part detectors learned by boosting are combined to form a joint likelihood of multiple inter-occluded humans. The detection problem is formulated as maximum posteriori (MAP) estimation. However, they inferred the depth of human from their image y-coordinates, the shape is approximated by 2-D ellipse and the visibility calculated according to the relative depth order of human objects.

Wang et al. [12] proposed a method for partial occlusion handling by combining HOG and local binary pattern features. To handle occlusion, they proposed a method which find the occluded part of the image using response of the HOG feature to the global detector. Then, the occlusion likelihood map is segmented by mean-shift approach. The segmented portion of the window with a majority of negative response is inferred as an occluded region. Part detectors are then applied on the non-occluded regions to achieve the final classification on the current scanning window. This method proposed a good idea to find occluded parts of the scanning window bu the final classification depends on part classifiers only is very week to give the final response.

Enzweiler et al. [4] presents a mixture-of-experts framework to classify pedestrians with partial occlusion. They trained a set of component-based expert classifiers on features derived from intensity, depth and motion. Occlusion are handled by computing expert weights based on visibility estimation from depth and motion information. The final result achieved by the combination of the part-based expert classifiers depends on the quality of the estimated weights. This estimation is inaccurate and depends on the scale of the parts.

Generic detectors assumes that pedestrians are fully visible, and their performance degrades when pedestrians are partially occluded. For example, many part-based models [7] summed the scores of part detectors. Ouyang et al. [15] handle the imperfectness of part detector by proposing a deformable part-based models to obtain accurate scores. The visibility of parts are modelled as hidden variables. Discriminative deep model is used to learn the relationships among overlapping parts.

Recently, Antunez et al. [16] encoded the visual structure of the objects by a 2D combinatorial map and a combinatorial pyramid. Searching for objects is performed as an error-tolerant submap isomorphism conducted at different levels of pyramid. The spatial relationships among the individual parts may not be taken into account. When these relationships are not considered, the descriptive ability is severely limited and many false negatives are highly confused with human object exploiting global and local shape descriptors avoids this problem. Among the techniques proposed to model the relations between the different parts of the object, such as star-structured graphical models [7].

The aforementioned methods did not utilize any occluded pedestrian patterns in the learning process. However, using occluded patterns in the final learning stage (after learning part-based classifiers) may leads to significant improvement in the detection. In this paper, we investigate the effect of adding occluded pattern on the final detection stage.

III. Proposed System for Occlusion-invariant Pedestrian Classification

The proposed system for partially occluded pedestrian detection system is shown in Fig. 1. The proposed system contains two phases, training phase (off-line) and detection phase (on-line). During training phase, a set of labelled training data contains positive and negative samples are collected to train multiple SVM classifiers. Each classifier is trained on a specific part of human body. The part based classifiers are trained using non-occluded pedestrian data only. After training part-based classifiers, the scores of both non-occluded and the artificially generated occluded pedestrians are used to create the new training vectors for the next stage. The collected scores are used to train restricted Boltzmann machine [17] using gradient-based contrastive algorithm. The final score is computed by adding the output of all RBM hidden neurons.

In the on-line detection phase, the input image is scanned using a sliding window of fixed size at dense positions and scales. Each detection window is divided into subregions and for each subregions, the response of the corresponding part-based classifier is computed. The scores are then passed to the trained RBM to estimate the final score of the window.

A. Part-based Classifiers

For pedestrian classification, the goal is to determine a class label for an unseen example x_i. We consider a two-class problem with classes C_0 (pedestrian) and C_1 (non-pedestrian). Since $P(C_1|x_i) = 1 - P(C_0|x_i)$, it is sufficient to compute the posterior probability $P(C_0|x_i)$ of the unseen pedestrian sample x_i.

The posterior probability $P(C_0|x_i)$ is approximated using a component-based mixture of experts model. A sample x_i is composed out of a set of overlapped K components which are semantically related to various human body parts. The final decision results is a weighted combination of scores which denoted as $P(C_0|x_i^k), k = 1, 2, ..., K$.

$$P(C_0|x_i) = W * P(C_0|x_i^k) \quad (1)$$

978-1-4799-2914-613/$31.00 ©2013 IEEE

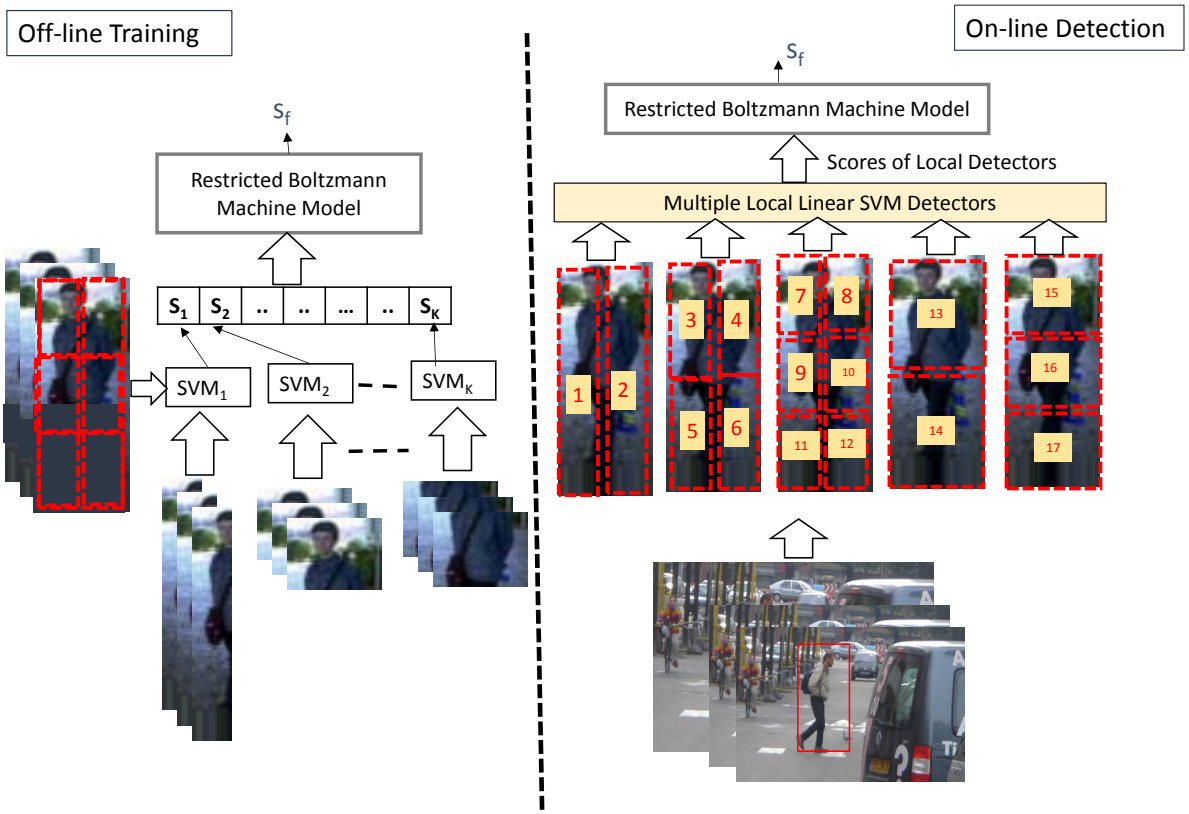

Fig. 1. Overview of the proposed occlusion invariant pedestrian detection system

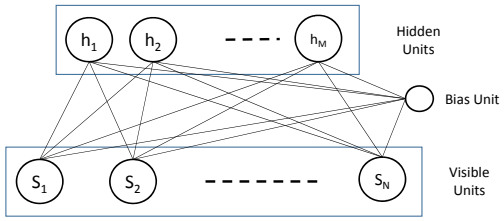

Fig. 2. An example of Restricted Boltzmann Machine Model

Note that the weight matrix W which combine the component classifier is learned from data. These weights allow to incorporate a model of partial occlusion into our system. The score probability for each parts is estimated from response of linear support vector machine classifier trained on a set of non-occluded pedestrians and negative patterns. Probability is estimated by applying method discussed in [18].

B. The Restricted Boltzmann Machine

The weight matrix W of the restricted Boltzmann machine model used to combine component features is learned from a set of both occluded and non-occluded pedestrian. A restricted Boltzmann machine [19] is a generative stochastic neural network which model a density over observed variables (i.e. over scores of part classifiers) that uses a set of hidden variables (represent presence of pedestrian). In Fig. 2, a diagram shows the structure of RBM is drawn. In this paper we associate the RBM's observed variables with scores of component classifiers and hidden variables with the presence of pedestrian. The final score is computed as a summation of the output of all hidden units.

A key characteristic of the RBM is that its stochastic hidden units are conditionally independent given the observed data. This property makes each hidden unit an independent estimator for pedestrian. The probability of observed variables in an RBM with parameter set θ is defined according to a joint energy of visible and hidden units $E(v, h; \theta)$, as a Gibbs distribution

$$p(v;\theta) = \frac{1}{Z(\theta)} \sum_{h} e^{-E(v,h;\theta)} \tag{2}$$

where v and h denote vectors of visible and hidden variables, and $Z(\theta)$ is the normalization constant. Commonly RBMs refer to a model with binary hidden and binary visible random variable. In binary RBM, the scores of component classifiers are first thresholded to give binary scores. The energy function of the binary RBM are defined as E_1 and E_2 respectively,

$$E_1(v, h; \theta) = -\sum_{i,j} v_i w_{ij} h_j - \sum_{i} b_i v_i - \sum_{j} c_j h_j. \tag{3}$$

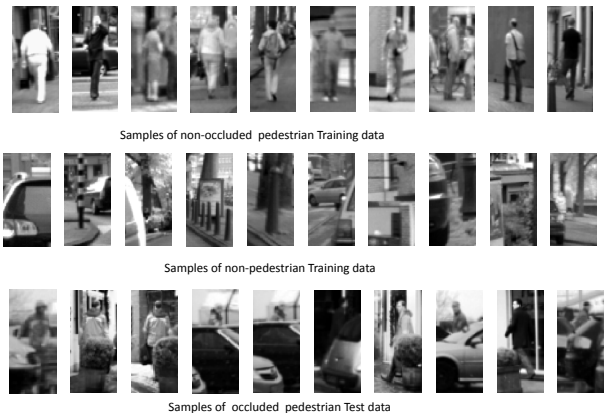

Fig. 3. Sample images from training and test data for Daimler data set

$$E_2(v,h,\theta) = E_1(v,h,\theta) + \frac{1}{2}\sum_i v_i^2 \tag{4}$$

where variables i and j iterate over observed and hidden units respectively, and $\theta = \{W, b, c\}$ is the model parameter set. The matrix W determines the symmetric interaction between pairs of hidden and visible units, and parameters b and c are bias terms that set the unary energy of the variables. Inference in RBMs is straightforward. In the binary RBM, conditionals are of the form

$$p(h_j = 1|v) = \sigma(c_j + \sum_i v_i w_{ij}), \tag{5}$$

$$p(v_i = 1|h) = \sigma(b_i + \sum_j h_j w_{ij}), \tag{6}$$

where $\sigma(x) = 1/(1+e^{-x})$ is the logistic sigmoid function. Typically, parameters of RBM are learned by maximizing the likelihood in a gradient ascent procedure. The gradient of the log-likelihood for an energy-based model is

$$\frac{\partial}{\partial\theta} = -\left\langle \frac{\partial E(v;\theta)}{\partial\theta} \right\rangle_{data} + \left\langle \frac{\partial E(v;\theta)}{\partial\theta} \right\rangle_{model} \tag{7}$$

where $E(v;\theta)$ is the free energy of v, and $\langle . \rangle_{data}$ and $\langle . \rangle_{model}$ denote expected value over all possible visible vectors v with respect to the data and model distribution. For an RBM $E(v;\theta) = -\log\sum_h e^{-E(v,h;\theta)}$ and $\partial E(v;\theta)/\partial\theta = \sum_h p(h|v;\theta)\partial E(v,h;\theta)/\partial\theta$.

Unfortunately computing expected value regarding an RBM distribution involves an exponential number of terms, which makes it intractable. However, Hinton [17] proposed an other objective function called *contrastive divergence* (CD) that can be efficiently minimized during training as an approximation to maximizing the likelihood.

In this paper, each hidden unit represent the estimation of pedestrian at different occlusion condition. The final score is computed as the summation over all hidden unit outputs.

IV. Experimental Results

The efficiency of the proposed system for occlusion invariant pedestrian detection was tested using Dailmer pedestrian dataset [4]. Dailmer dataset is chosen hear because it contains labelled real partially occluded pedestrian. Our approach is evaluated in a pedestrian classification setting, where we assume that initial pedestrian location hypotheses already given, e.g. using methods described in [5]. In our experiments, we focus only on the central part of a pedestrian detection system, i.e. the classifier, to eliminate auxiliary effects arising from various detector parameters such as grid granularity, non-maximum suppression, scene and processing constraints or tracking.

A. Daimler Occluded Pedestrian Dataset

Daimler data set [4] contains a set of manually labelled pedestrian and non-pedestrian bounding boxes in images captured from a vehicle-mounted calibrated stereo camera rig in an urban environment. For each manually labelled pedestrian, a set of additional samples are generated by geometric jittering. Non-pedestrian samples were the result of a shape detection preprocessing step with relaxed threshold setting, i.e. containing a bias towards more "difficult" patterns. The Daimler data set consists of $52,112$ pedestrians and $32,465$ non-pedestrians training samples, the test data divided into partially occluded set and non-occluded test set. The partially occluded test set contains $11,160$ pedestrians and $16,253$ non-pedestrians, while non-occluded test set contains $25,608$ and $16,235$ pedestrians and non-pedestrians test images respectively. Example of images from the dataset are shown in Fig. 3.

B. Experimental Setup

Training and test samples have a resolution of 36×84 pixels with a $6-$pixel border around the pedestrians. In our experiments, the number of component classifiers K is set to 17 components, each component represent different region of the human body. The reason for choosing large number of components because we want to cover all possible occlusion conditions. Note that the components overlap with each other and represent a semantic information about pedestrian i.e. head, torso, legs, right hand, left hand, right leg, left leg, right-human body parts, left-human body parts,..,etc. This redundancy is required to improve the performance. The number of hidden neurons used for RBM model is chosen experimentally, we select 3 hidden neurons only to estimate the final pedestrian score at various occlusion conditions.

HOG features [6] are chosen among other different features to component expert linear SVM classifiers. That is because HOG features are still among the best performing feature set available. Similar to [13], we compute histograms of oriented gradients with 12 orientation bins and 6×6 pixel cells, accumulated to overlapping 12×12 pixel blocks with a spatial shift of 6 pixels. For classification, we employ linear support vector machines (SVMs).

To train component classifiers, only non-occluded pedestrians (and non-pedestrians samples) are used. However, to train RBM model, we augment the previously used training data with an artificially generated occluded pedestrian data.

For testing, we evaluate performance using two different test sets: one involving non-occluded pedestrians and

978-1-4799-2914-613/$31.00 ©2013 IEEE

Fig. 4. Sample images of artificially occluded generated pedestrian using various occlusion sizes

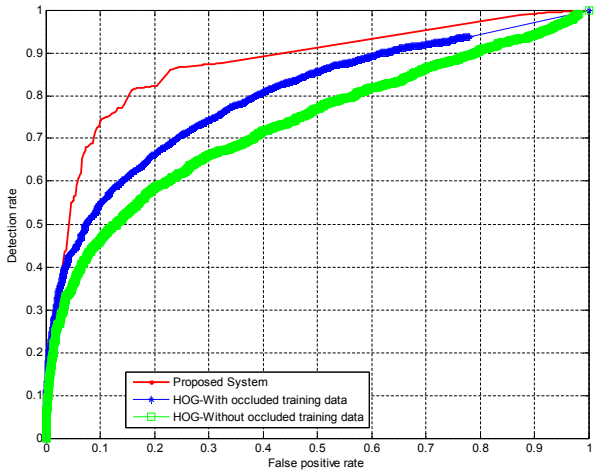

Fig. 5. Classification performance of partially occluded test data

one contains real partially occluded pedestrians. The non-pedestrian samples are same for both test sets.

C. Occluded Data Set Generation

In this work, a set of artificial occluded pedestrian samples are generated. These samples are used to train RBM probabilistic model to estimate the final score of the detection window. The generated data assumes that the lower part of the pedestrian is the most probable occluded part. This assumption used to randomly occlude various areas of the lower human body part. The occlusion pattern is created from the contextual pixels around the pedestrian. Various occluded pattern sizes with diffident aspect ratios are used in the experiments. Occluded patterns of sizes $\{6 \times 6, 6 \times 12, 12 \times 6, 12 \times 12, ...\}$ pixel are used to generate various partially occluded pedestrians samples. Fig. 4 shows examples of the generated samples.

D. Performance on Partially Occluded Test Data

In this experiment, we evaluate the performance of the proposed method to handle occlusion. The base-line classifier used in the comparison is full-body HOG approach of [6] trained on non-occluded pedestrian data set. In order to investigate the effect of augmenting training data with occluded patterns, another full-body HOG classifier is trained using the set of artificially generated occluded pedestrians. Our proposed system is evaluated after training the RBM model with the artificially occluded generated data. Results in terms of ROC performance are given in Fig. 5. The results reveal that the original full-body HOG approach deteriorate when test samples are occluded. However, the performance of the full-body HOG approach trained on both

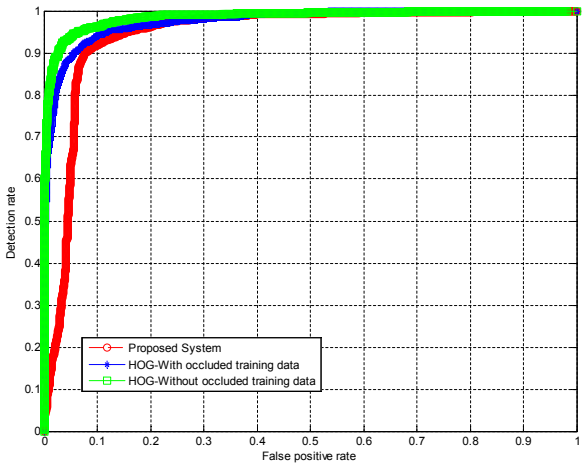

Fig. 6. Classification performance of non-occluded test data

occluded and non-occluded samples sightly increases its performance. Our proposed part-based classifier combined with RBM model outperform both methods. The results prove show that augmenting the non-occluded pedestrian training data with occluded samples improve the detection rate on occluded samples. However part-based classifier method is more robust than full-body based classifier.

E. Performance on Non-occluded Test Data

After demonstrating significant performance boosts on partially occluded test data, we evaluate the performance of the proposed system using non-occluded pedestrians (and non-pedestrians) as test set. Fig. 6 shows the performance of the proposed method compared with full-body HOG classifier trained on occluded and non-occluded data set. The best performance is achieved by the full-body HOG classifier without occluded training data. However, the performance of our proposed method is still comparable. The results also show that in case of non-occluded pedestrian data set, the performance of full-body classifier is better than part-based classifiers. Furthermore, augmenting training data with the artificially generated occluded samples further deteriorate the performance of both method.

V. CONCLUSIONS

The paper presented a new occlusion invariant pedestrian classification system based on the combination of component classifiers and restricted Boltzmann machine mode. For partially occluded dataset, we obtained an improvement versus full-body HOG approach. The performance of full-body classifier in the occlusion test data is improved when adding occluded samples in the training data. Using artificially occluded generated data in the final training stage of our proposed method highly improve the detection rate of real occluded data. For the non-occluded dataset, occlusion handling does not appreciably deteriorate results.

VI. ACKNOWLEDGEMENTS

Parts of this research were supported by Egyptian government, MEXT, Grant-in-Aid for Scientific Research and

Grant-in-Aid for Young Scientists.

REFERENCES

[1] T. Gandhi and M. Trivedi, "Pedestrian protection systems: Issues, survey, and challenges," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 8, no. 3, pp. 413–430, 2007.

[2] D. Geronimo, A. M. Lopez, A. D. Sappa, and T. Graf, "Survey of pedestrian detection for advanced driver assistance systems," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 7, pp. 1239–1258, 2010.

[3] D. M. Gavrila and S. Munder, "Multi-cue pedestrian detection and tracking from a moving vehicle," *International journal of computer vision*, vol. 73, no. 1, pp. 41–59, 2007.

[4] M. Enzweiler and D. M. Gavrila, "Monocular pedestrian detection: Survey and experiments," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 12, pp. 2179–2195, 2009.

[5] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 4, pp. 743–761, 2012.

[6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, pp. 886–893, IEEE, 2005.

[7] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 9, pp. 1627–1645, 2010.

[8] S. Maji, A. C. Berg, and J. Malik, "Classification using intersection kernel support vector machines is efficient," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pp. 1–8, IEEE, 2008.

[9] P. Dollar, P.r, Z. Tu, P. Perona, and S. Belongie, "Integral channel features," in *British machine vision conference*, pp. 1–11, 2009.

[10] P. Dollar, P.r, S. Belongie, and P. Perona, "The fastest pedestrian detector in the west," *BMVC 2010, Aberystwyth, UK*, 2010.

[11] P. Viola, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," *International Journal of Computer Vision*, vol. 63, no. 2, pp. 153–161, 2005.

[12] X. Wang, T. X. Han, and S. Yan, "An hog-lbp human detector with partial occlusion handling," in *Computer Vision, 2009 IEEE 12th International Conference on*, pp. 32–39, IEEE, 2009.

[13] M. Enzweiler, A. Eigenstetter, B. Schiele, and D. M. Gavrila, "Multi-cue pedestrian classification with partial occlusion handling," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pp. 990–997, IEEE, 2010.

[14] B. Wu and R. Nevatia, "Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, vol. 1, pp. 90–97, IEEE, 2005.

[15] W. Ouyang and X. Wang, "A discriminative deep model for pedestrian detection with occlusion handling," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 3258–3265, IEEE, 2012.

[16] E. AntAnez, R. Marfil, J. P. Bandera, and A. Bandera, "Part-based object detection into a hierarchy of image segmentations combining color and topology," *Pattern Recognition Letters*, no. 0, pp. –, 2013.

[17] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural computation*, vol. 14, no. 8, pp. 1771–1800, 2002.

[18] J. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Advances in large margin classifiers*, vol. 10, no. 3, pp. 61–74, 1999.

[19] R. Salakhutdinov, A. Mnih, and G. Hinton, "Restricted boltzmann machines for collaborative filtering," in *Proceedings of the 24th international conference on Machine learning*, pp. 791–798, ACM, 2007.