

Motorcycles detection using Haar-like features and Support Vector Machine on CCTV camera image

Imam Teguh Mulyawan^{1,*}, Adhi Prahara²

Informatics Department, Universitas Ahmad Dahlan

Jalan Ring Road Selatan, Tamanan, Banguntapan, Bantul, Yogyakarta 55166, Indonesia

¹ *imamteguh1400018200@webmail.uad.ac.id*

² *adhi.prahara@tif.uad.ac.id*

** corresponding author*

ABSTRACT

Traffic monitoring system allows operators to monitor and analyze each traffic point via CCTV camera. However, it is difficult to monitor each traffic point all the time. This problem leads to the development of intelligent traffic monitoring system using computer vision technology which one of the features is vehicle detection. Vehicle detection still poses a challenge especially when dealing with motorcycles that occupy the majority of the road in Indonesia. In this research, a motorcycle detection method using Haar-like features and Support Vector Machine (SVM) on CCTV camera image is proposed. A set of preprocessing procedure is performed on the input image before Haar-like features extraction. The features then classified using trained SVM model via sliding window technique to detect motorcycles. The test result shows 0.0 log average miss rate and 0.9 average precision. From the low miss rate and high precision, the proposed method shows promising solution in detecting motorcycle from CCTV camera image.

Keywords:

Motorcycles detection
Object detection
Computer vision
Haar-like features
SVM

I. Introduction

Traffic monitoring is an important task especially in big cities that have large volume of vehicle passing through their road every day. Traffic monitoring system allows operators to monitor and analyze each traffic point via closed circuit television (CCTV) camera [1]. CCTV cameras can be placed in the roadsides or driveways [2]. From the CCTV cameras, we can monitor traffic congestion, traffic violation and count how many vehicle is passing in that road at a time to perform traffic management and control. However, it is difficult to monitor each traffic point all the time. This problem leads to the development of intelligent traffic monitoring system using computer vision technology [3], [4].

In intelligent traffic surveillance system, many objects need to be detected and recognized e.g. vehicle, road sign, and people. Vehicle detection and recognition is the most important and challenging stage of traffic surveillance using computer vision techniques due to the variability of on-road driving environments [4] and various poses of vehicle produced from the different viewpoint of the CCTV camera [5], [6]. Many researchers have been developing methods to handle this problem which generally consists of image acquisition, preprocessing, feature extraction, and classification step. A preprocessing step is performed to prepare the input image or video captured from CCTV camera before the feature extraction procedure. The popular features in vehicle detection is Haar-like features [7]–[11] and Histogram of Oriented Gradients (HOG) descriptors that perform well in vehicle detection [2], [5], [6], [12]–[14]. Haar-like features are suitable for vehicle detection because it forms a compact representation, encode edge and structural information, capture information from multiple scales, and especially can be computed efficiently [7].

Once the feature extracted, the detection method applies a classifier such as Support Vector Machine (SVM) which has been widely used in object detection and recognition problems e.g. to detect human/pedestrian, face, fruits, vehicle, number plate, and so on [5]–[7], [13]–[18]. Prahara *et al.* proposed a method to detect car using HOG and SVM by estimating the road direction [5]. The idea uses four categories of road direction to determine the pose of a car and choose the right



detector. By doing that, the proposed method can handle different pose of cars from various viewpoint of CCTV camera. Bougharriou *et al.* also proposed a car detection method using HOG and linear SVM and achieves robustness and good precision from various scene [13]. Wen *et al.* proposed vehicle detection using Haar-like features and perform rapid and effective feature selection via AdaBoost. An improved normalization for the selected features is used to reduce the intra-class difference and increase the inter-class variability. The result shows a speed up in the feature selection process and better detection performance than the state-of-the-art methods [7]. Haselhoff and Kummert proposed a vehicle detection method using Haar and triangle features which computed based on four integral images. The proposed method is based on boosted cascaded classifiers, Haar and triangle features, and adaptive sliding window and kalman filter to locate and track the vehicles.

This research proposes motorcycle detection method using Haar-like features and SVM. The method receives input image captured from CCTV camera and perform preprocessing on the image before feature extraction. Haar-like features extracted from integral image then classified using SVM to generate the motorcycle detector. The detector then used to detect front-view motorcycles from the image. The rest of this paper is organized as follows. Section 2 presents the methodology, section 3 presents the result and discussion, and section 4 describes the conclusion of this work.

II. Methodology

This research proposed a method to detect motorcycles using Haar-like features and SVM from CCTV camera video. The method consists of preprocessing, feature extraction, and classification step. The general procedure of the proposed method is shown in Fig. 1 where the blue arrows represents the flow of the training step and the orange arrows represents the flow of the testing step.

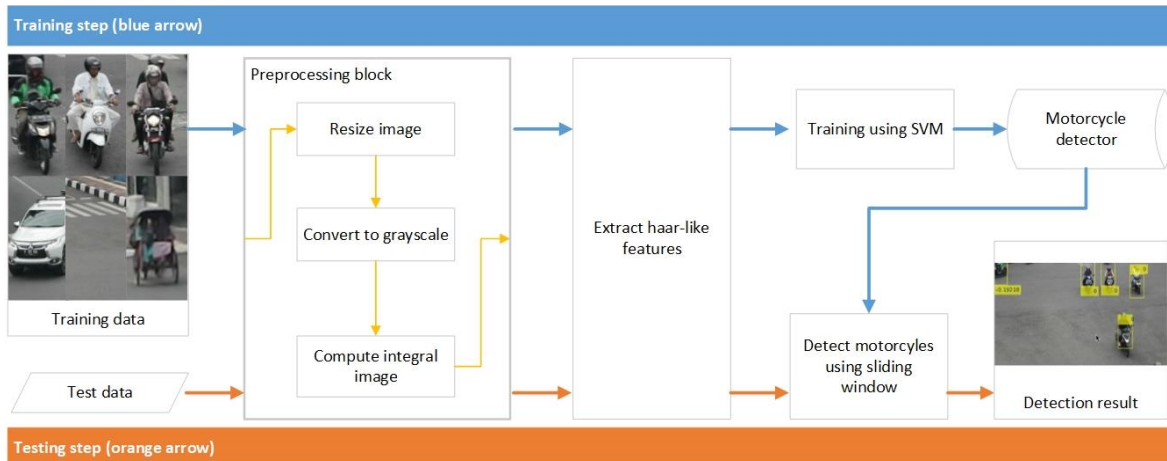


Fig. 1. General procedure of the proposed motorcycle detection method.

The details explanation of each step is described as follows.

A. Preprocessing

CCTV camera allows HD video resolution streamed through the network. The resolution can be resized to a smaller resolution to reduce the number of pixel to be processed in the feature extraction step. The next step is grayscale conversion to reduce the image dimension from three channels RGB to single channel grayscale. The grayscale image range is 0-255 (black to white). To improve the speed of Haar-like features extraction, integral image will be used. Integral image is an image that each of its pixel is a sum of the left-top to the right-bottom rectangular area from that pixel location [19]. Generally, integral image has principle of weighted addition where the weight is the pixel values that added to the original image. The illustration of integral image from original image is shown in Fig. 2.

1	2	3
4	5	6
7	8	9

1	3	6
5	12	21
12	27	45

(a) Original image (b) Integral image

Fig. 2. The example of integral image from the original image.

By getting the pixel value of integral image, the sum of all pixels in the rectangular region can be computed with only four values. These four pixels denoted by A, B, C, D are located on each corner of the rectangular region from the corresponding pixel in integral image as shown in Fig. 3. The computation is shown in (1).

$$\text{sum} = D - B - C + A \tag{1}$$

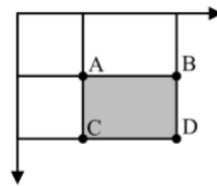


Fig. 3. The four values in integral image to compute the sum of all pixels in the rectangular region.

B. Feature extraction

Feature extraction using Haar-like features can be found in many object detection methods. Haar refers to the mathematical function (Haar wavelet) which has rectangular shape. Haar-like features becomes popular after introduced by Viola and Jones in their face detection research [20]. Viola-Jones version of Haar-like features consists of edge features, line features, and four-rectangle features as shown in Fig. 4 where the rectangular region divided into white and black region. Haar-like features place the filters illustrated in Fig. 4 on the image.

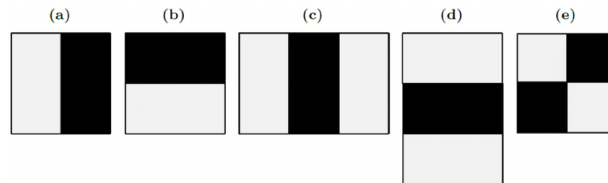


Fig. 4. Haar-like features models.

The difference between sum of the pixels in the white region and sum of the pixels in the black region will be used as a threshold. This computation can be simplified using integral image. The Summed Area Table or SAT(x, y) is the sum of the pixels from the upper-left corner (0, 0) to the lower-right corner (x, y) of the rectangle and can be computed using (2) [21].

$$SAT(x, y) = \sum_{x' \leq x, y' \leq y} I(x', y') \tag{2}$$

where, I(x', y') is a pixel value on the original image. It can be calculated by all pixels from left to right and from top to bottom using (3) and (4).

$$SAT(x, y) = SAT + SAT(x - 1, y) + I(x, y) - SAT(x - 1, y - 1) \tag{3}$$

$$SAT(-1, y) = SAT(x, -1) = SAT(-1, -1) = 0 \tag{4}$$

A positive rectangle of pixels $r = (x, y, w, h, 0)$ can be computed by looking up the integral image using (5).

$$\begin{aligned} RecSum = & SAT(x - 1, y - 1) + SAT(x + w - 1, y + h - 1) \\ & - SAT(x - 1, y + h - 1) - SAT(x + w - 1, y - 1) \end{aligned} \tag{5}$$

C. Classification using Support Vector Machine (SVM)

The features are classified using Support Vector Machine (SVM) into two categories: motorcycle or background. The concept of SVM classification originated from the two classes that requires

training data in positive and negative samples. SVM tries to find the best hyperplane to separate two classes and maximizes the margin between the two classes. The decision function of SVM is shown in (6) and (7) [22]. In the training step, if there are given training data $x_i \in R^n, i = 1, \dots, l$ in two classes and label $y \in R^l$ such that $y_i \in \{1, -1\}$, it can be solved using (6).

$$\min_{w, b, \xi} \frac{1}{2} w^T w + C \sum_{i=1}^l \xi_i \quad (6)$$

$$\text{subject to } y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, \dots, l$$

where w is weight vector, b is bias, ξ_i is slack variables, $\phi(x_i)$ maps x_i into higher dimensional space and $C > 0$ is the regularization parameter and the decision function is shown in (7).

$$\text{sgn}(w^T \phi(x) + b) = \text{sgn}(\sum_{i=1}^l y_i \alpha_i K(x_i, x) + b) \quad (7)$$

where $K(x_i, x_j)$ is the kernel function. After training process, parameter $y_i \alpha_i \forall i$, b , label names, support vectors, and kernel parameter saved as trained SVM model.

In classification step, voting strategy is performed for each data x which will be designated to be in a class with the maximum votes. After training process we will get variable w , x , and b for each class, then the classification process can be done with these steps:

1. Calculate kernel.
2. Calculate decision function using (7).
3. Repeat step 1 and 2 for other classes.
4. Determine the class by function which gives the most maximum result.

Optimal parameter can be selected using k -fold cross validation which is a method to do cross validation by dividing training data into k set which has $(k-1)$ as training data and the rest will be the test data.

D. Non maximum suppression

The result from SVM motorcycle detector maybe overlapped because the use of sliding window technique. These overlapped results can be suppressed into one result using non-maximum suppression (NMS). It combines the detections of the same class together into one object if the detections overlap. An illustration of NMS is shown in Fig. 5 where the red rectangles are the prediction results and the green rectangles are the result of NMS grouping those red rectangles.

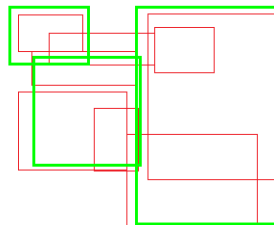


Fig. 5. The illustration of non-maximum suppression.

The three green regions are selected because they are the highest probability regions in the area that do not overlap at a fraction greater than α with a higher probability region. The red regions are suppressed because the regions they occupy were already occupied at a fraction greater than α by a higher probability region.

E. Evaluation metrics

The performance of object detection method can be evaluated using miss-rate and precision. Miss-rate measures the rate of undetected objects by the system while precision measures how precise the system in detecting objects. The evaluation compares the ground-truth object annotation in a region of rectangle by human and the prediction made by the system. The performance evaluation graph of log average miss rate plots the miss rate against FPPI (false positive per image) while the graph of precision plots the precision against recall.

III. Result and Discussion

The proposed motorcycle detection method written using Matlab and runs on laptop with Intel i5 processor and 16GB of RAM. The number of training data is 2.034 images consists of 856 positive samples (motorcycle images) and 1.171 negative samples (background images). The motorcycle data captured in front-view with size of 150x300 pixels. Fig. 6 shows the positive and negative samples used in this research.

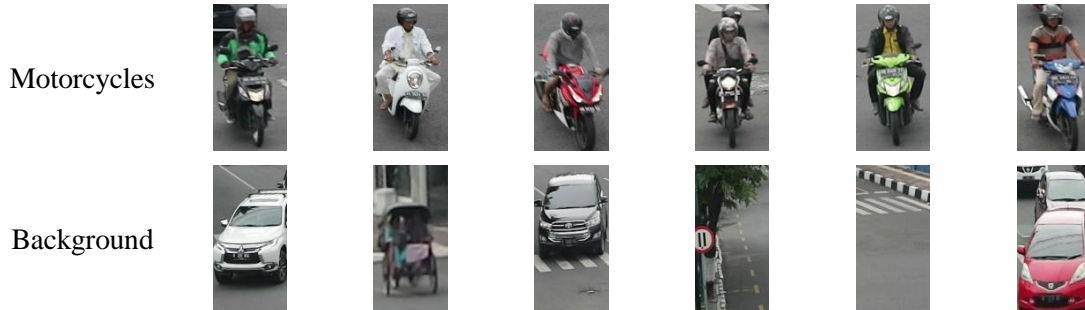


Fig. 6. Sample of the training dataset.

A. Motorcycle detection result

The trained motorcycle detector is used with sliding window technique to detect motorcycles in the image. Some preprocessing steps are applied to reduce the computation time. In the feature extraction step using Haar-like features, smaller image size and grayscale color space are used to reduce the number of pixels to be processed. Integral image is used to fasten the feature extraction procedure. The trained motorcycle detector is applied in each step of sliding window which has size of 30x58 pixels. The test conducted on some frames captured from three CCTV videos with 426x240 pixels resolution. Table 1 shows a sample of the detection result.

The detector returns a confidence score when detecting an object. The higher confidence score shows that the proposed method is confidence that the object is a motorcycle and otherwise. The proposed method using threshold of -0.2 to filtered out the predicted result. If the confidence score is higher than -0.2 then the object is classified as motorcycle, otherwise it is a background. There are some misclassification and overlapped detection result on sample no. 2, 3 and 5. The misclassification caused by the scaling of sliding window that did not cover the whole motorcycle (see sample no. 3 and 5, the detector detects one motorcycle as two motorcycles). For this case, the detector still predicted right because the predicted area is a motorcycle. The detector also detects the road between two motorcycles as motorcycles because some parts of the motorcycles covered in the sliding window.

Table 1. The example of motorcycle detection result.

No	Sample	Actual	Predicted	Confidence Score Per Object
1		5	5	-0.19218, 0, 0, 0, 0
2		4	6	-0.87134, 0, -0.48245, 0, 0, 0
3		6	9	0, 0, 0, 0, -0.18182, 0, 0, -0.15644, 0

4		4	4	0, 0, 0, 0
5		7	8	0, 0, 0, 0, 0, 0, 0

B. Performance evaluation

Log average miss rate and precision are used to measure the performance of the proposed motorcycle detection. The ground-truth rectangle annotation of objects is compared with the predicted rectangle result. If the overlapped bounding box is greater than threshold, then it is counted as true positive. The performance evaluation results are shown in Fig. 7 where the log average miss rate shows 0.0 score and the average precision shows 0.9 score. Low miss rate score means the proposed method can detect almost all of the motorcycle objects in the image. High precision score means the proposed method precisely predicts the detected object which is actually motorcycle as motorcycle. The low miss rate and high precision score means that the proposed method performs well in the motorcycle detection problem.

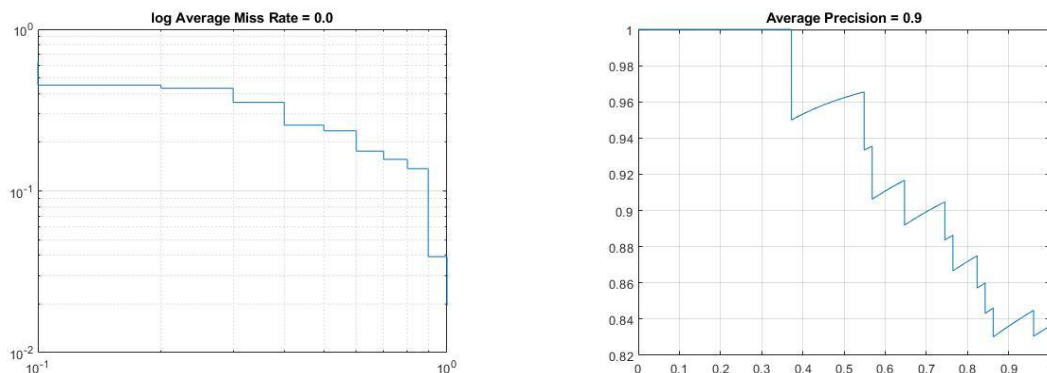


Fig. 7. The performance evaluation result.

IV. Conclusion

This research proposed motorcycle detection using Haar-like features and Support Vector Machine on CCTV camera image. Based on the experiment, the proposed method achieves low miss rate and high precision when detecting motorcycle. A low miss rate means that the proposed method able to detect most of the objects in the image. This will be complemented with a high precision that means all of the detected objects are indeed motorcycles. For the future works, more motorcycles data will be added so the detector is able to detect not only front-view but also side-view motorcycles and will be further improved to achieve real time detection.

References

- [1] A. Nurhadiyatna, B. Hardjono, A. Wibisono, W. Jatmiko, and P. Mursanto, "ITS information source: Vehicle speed measurement using camera as sensor," in *2012 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 2012, pp. 179–184.
- [2] K. B. Saran and G. Sreelekha, "Traffic video surveillance: Vehicle detection and classification," in *2015 International Conference on Control, Communication and*

- Computing India, ICCV 2015*, Mar. 2016, pp. 516–521, doi: 10.1109/ICCC.2015.7432948.
- [3] S. Messelodi, C. M. Modena, and M. Zanin, “A computer vision system for the detection and classification of vehicles at urban road intersections,” *Pattern Anal. Appl.*, vol. 8, no. 1–2, pp. 17–31, Sep. 2005, doi: 10.1007/s10044-004-0239-9.
- [4] Z. Yang and L. S. C. Pun-Cheng, “Vehicle detection in intelligent transportation systems and its applications under varying environments: A review,” *Image and Vision Computing*, vol. 69. Elsevier Ltd, pp. 143–154, Jan. 01, 2018, doi: 10.1016/j.imavis.2017.09.008.
- [5] A. Prahara and Murinto, “Car detection based on road direction on traffic surveillance image,” in *Proceeding - 2016 2nd International Conference on Science in Information Technology, ICSITech 2016: Information Science for Green Society and Environment*, Feb. 2017, pp. 344–349, doi: 10.1109/ICSITech.2016.7852660.
- [6] A. Prahara, A. Azhari, and Murinto, “Vehicle pose estimation for vehicle detection and tracking based on road direction,” *Int. J. Adv. Intell. Informatics*, vol. 3, no. 1, pp. 35–46, Mar. 2017, doi: 10.26555/ijain.v3i1.88.
- [7] X. Wen, L. Shao, W. Fang, and Y. Xue, “Efficient feature selection and classification for vehicle detection,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 3, pp. 508–517, Mar. 2015, doi: 10.1109/TCSVT.2014.2358031.
- [8] Y. Tang, C. Zhang, R. Gu, P. Li, and B. Yang, “Vehicle detection and recognition for intelligent traffic surveillance system,” *Multimed. Tools Appl.*, vol. 76, no. 4, pp. 5817–5832, Feb. 2017, doi: 10.1007/s11042-015-2520-x.
- [9] S. M. Elkerdawi, R. Sayed, and M. ElHelw, “Real-Time Vehicle Detection and Tracking Using Haar-Like Features and Compressive Tracking,” Springer International Publishing, 2014, pp. 381–390.
- [10] D. Chen, G. Jin, L. Lu, L. Tan, and W. Wei, “Infrared Image Vehicle Detection Based on Haar-like Feature,” in *Proceedings of 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference, IAEAC 2018*, Dec. 2018, pp. 662–667, doi: 10.1109/IAEAC.2018.8577211.
- [11] A. Haselhoff and A. Kummert, “A vehicle detection system based on haar and triangle features,” in *IEEE Intelligent Vehicles Symposium, Proceedings*, 2009, pp. 261–266, doi: 10.1109/IVS.2009.5164288.
- [12] S. Madhogaria, P. Baggenstoss, M. Schikora, W. Koch, and D. Cremers, “Car detection by fusion of HOG and causal MRF,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 1, pp. 575–590, Jan. 2015, doi: 10.1109/TAES.2014.120141.
- [13] S. Bougharriou, F. Hamdaoui, and A. Mtibaa, “Linear SVM classifier based HOG car detection,” in *2017 18th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering, STA 2017 - Proceedings*, Mar. 2018, vol. 2018-January, pp. 241–245, doi: 10.1109/STA.2017.8314922.
- [14] Y. Xu, G. Yu, Y. Wang, X. Wu, and Y. Ma, “A Hybrid Vehicle Detection Method Based on Viola-Jones and HOG + SVM from UAV Images,” *Sensors*, vol. 16, no. 8, p. 1325, Aug. 2016, doi: 10.3390/s16081325.
- [15] A. Prahara, A. Pranolo, and R. Drezewski, “GPU Accelerated Number Plate Localization in Crowded Situation,” *Int. J. Adv. Intell. Informatics*, vol. 1, no. 3, pp. 150–157, 2015, [Online]. Available: <http://ijain.org/index.php/IJAIN/article/view/46>.
- [16] G. Guo, S. Z. Li, and K. Chan, “Face recognition by support vector machines,” in *Proceedings - 4th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2000*, 2000, pp. 196–201, doi: 10.1109/AFGR.2000.840634.
- [17] Y. Zhang and L. Wu, “Classification of Fruits Using Computer Vision and a Multiclass Support Vector Machine,” *Sensors*, vol. 12, no. 9, pp. 12489–12505, Sep. 2012, doi: 10.3390/s120912489.
- [18] D. I. R. González and J.-B. Hayet, “Fast Human Detection in RGB-D Images with Progressive SVM-Classification,” in *Image and Video Technology: 6th Pacific-Rim Symposium, PSIVT 2013, Guanajuato, Mexico, October 28-November 1, 2013. Proceedings*, R. Klette, M. Rivera, and S. Satoh, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 337–348.
- [19] C. Messom and A. Barczak, “Fast and efficient rotated haar-like features using rotated

- integral images,” in *Australian Conference on Robotics and Automation*, 2006, pp. 1–6.
- [20] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, pp. I-511-I-518, doi: 10.1109/CVPR.2001.990517.
- [21] D. Chen, G. Jin, L. Lu, L. Tan, and W. Wei, “Infrared Image Vehicle Detection Based on Haar-like Feature,” *Proc. 2018 IEEE 3rd Adv. Inf. Technol. Electron. Autom. Control Conf. IAEAC 2018*, no. Iaeac, pp. 662–667, 2018, doi: 10.1109/IAEAC.2018.8577211.
- [22] C. C. Chang and C. J. Lin, “LIBSVM: A Library for support vector machines,” *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, Apr. 2011, doi: 10.1145/1961189.1961199.