Automatic and Invariant Recognition of Traffic Sign with Color-HMAX and Deep ELM Based Classifier

Khan Babar[1], Wang Zhijie[1], Muhammad Anwar Ahmed[2], Han Fang[1] and Zain Anwar Ali[3]

[1]*College of Information Science and Technology, Donghua University, Shanghai, P. R. China.*
[2]*Electronic Engineering Department, Sir Syed University of Engineering & Technology, Karachi, Pakistan.*
[3]*College of Automation Engineering, Nanjing University of Aeronautics & Astronautics, Nanjing, Jiangsu, China.*
*413039@mail.dhu.edu.cn, wangzj@dhu.edu.cn,
engr_anwar_ahmed@outlook.com, yadiahan@163.com,
zainanwar86@hotmail.com.*

Abstract

One of the several challenges in the development of an automatic traffic sign recognition system is the invariant recognition of various traffic signs irrespective of the occlusion and changes in position, size, orientation, illumination and viewpoint. An automatic traffic sign recognition system usually consists of two constituent parts, namely features extractor and features classifier. Most popular Traffic Sign Recognition (TSR) systems usually comprises of a CNN (Convolutional Neural Network) based features extractor and classifier, which achieve significant results but at the cost of a very complex and huge network. Furthermore, since classical NN (Neural Network) based classifiers are trained by Gradient Descent for weights tuning, therefore their generalization capability is limited. To improve the performance further, in this paper we make use of the hierarchical feed-forward model HMAX for the feature extraction and deep extreme learning machine based classifier for the classification of the traffic signs. HMAX model generalizes well and facilitates the features extraction invariant to position, size, orientation, viewpoint and luminance. This robust feature representation is then fed to the deep ELM based feature classifier for the classification of the traffic signs to the respective class category. The recognition and classification accuracy of the developed algorithm was experimentally tested on German Traffic Sign Recognition Benchmark (GTSRB). Achievement of excellent recognition performance is demonstrated with improved generalized, invariant and robust feature representation.

Keywords: Traffic Sign Recognition, hierarchical feed-forward neural network, HMAX, deep extreme learning machine

1. Introduction

The recent advancement in the development of Artificial Intelligence using Advance Machine Learning Algorithms have given birth to a new era of technology, where traditional cars are being replaced by self-driving cars, human labor is being replaced by automated robots. This new era of technology has not only reshaped the traditional machines but has also achieved improvements in terms of faster computation speed, better accuracy, and cost effectiveness. The importance of a robust automatic traffic sign recognition system for applications like advance driver assistance system [1], for the development of self-driving cars [2] is now certain. Although a lot of algorithms have

been developed, however the major challenge remains to develop an automatic traffic sign recognition system with optimum accuracy and faster recognition speed irrespective of the occlusion and changes in the position, size, orientation, illumination and viewpoint. An automatic traffic sign recognition system mainly consists of two stages namely, Traffic Sign Extraction (TSE) and Traffic Sign Classification (TSC) stage. The responsibility of a TSE module is to filter the image and perform feature extraction to identify the region of interest (ROI) particularly the boundaries of the traffic sign from a given image. Whereas the TSC module is responsible to perform the feature matching for recognizing and classifying the traffic sign to their respective class category. A good TSE and TSC module should find all the traffic signs in an image while maintaining low false negative triggers.

Traditionally, traffic sign detection methods are divided into two main categories namely, shape based methods [3] and color based methods. Color based methods primarily fragments a given color image to provide a region of interest to process further. The major difficulty of using these methods is to assess the color information correctly from the image irrespective of illumination variance due to several weather conditions. Researchers use different color spaces and thresholds to segment the traffic signs from the given image. HIS color space is widely used to segment the traffic signs from the image due to their robustness to illumination variance and weather conditions [4]. However dela Escalera *et al*., Yalic *et al*., Ruta *et al*. [5] used the RGB color space with thresholds for segmentation. Khan *et al*. used CIELab and gabor filters to represent a color image because of independent control over intensity and color information [6]. On the other hand, Hough Transform (HT) is widely used among shape based methods. Although these HT based detection may offer satisfactory results but their main disadvantages are highly complex computations and huge storage requirements. Other methods utilized corners in an image to detect traffic signs, like the distance transform (DT) [4]. Although these methods detect certain shapes quite well, but they are time consuming so they are unsuitable for real time applications [7].

Most traffic sign detection methods described earlier either fall under color based methods or shape based methods [8]. In recent studies, researchers have combined the color based method with shape based method to achieve better results [9]. Lillo Castellano *et al*. [9] used HSI color space based segmentation along with machine learning to detect the traffic sign shape. In recent works, traffic sign detection methods were based on the local stability of the traffic sign regions rather than the boundary color [8, 10].

Many researchers have developed different machine learning algorithms based on neural network for the features classification by training the input weights as well as output weights of the network layers. However, Schmidt *et al*. [11] proposed a feed-forward neural network as a features classifier and experimentally proved that if the output weights of the network are well tuned then high accuracy can be achieved even without tuning the weights of most hidden layers. Guang-Bin-Huang *et al*. [12] proposed a powerful single layer feed-forward neural network based learning algorithm called Extreme Learning Machine (ELM) motivated by the work mentioned above [11]. In a recent work, deep ELM network was developed by using the stacked de-noising auto-encoders (SDA) to train the ELM with features to detect the ships from remote images [13]. The results showed that the deep ELM network outperformed support vector machine (SVM) and other competent classifiers in terms of accuracy as well as time consumption.

In this paper, we propose a biologically inspired method for traffic sign detection and recognition from color images by performing joint processing of the shape and color cues. For the traffic sign detection, we employ feature extraction based on HMAX [14] algorithm combined with the color opponent channels [15] for the joint processing of the color cues along with the shape features. Whereas for the recognition and classification,

we employ deep extreme learning machine (DELM) [12] to classify the extracted traffic sign features from the input image patch to the respective class category within preformed set of traffic signs. Finally we assess the performance of the proposed network quantitatively using German Traffic Sign Recognition Benchmark (GTSRB) [16]. In the following section, we first describe the outline of the proposed network framework with the help of a flowchart. Then we describe each module of the proposed network and their integration. In the later section, we present the results and our conclusions.

2. Method

2.1. Proposed Framework (TSD&C)

The overall workflow of the proposed network framework is summarized with the help of a flowchart in Fig 1. For a given input image in RGB (Red-Green-Blue) color, we first separate each color channel from the input image. We then combined these three color channels into three opponent color channels Red-Green, Yellow-Blue and White-Black. As proposed by the classical opponent color channel theory, these opponent channels are inspired by the color representation in visual cortex. These opponent channels are then processed further by the traffic sign feature extraction module based on HMAX model for machine vision to extract relevant features representation. The extracted features are then processed further by the traffic sign recognition module based on deep ELM (DELM) to classify the detected traffic sign to a respective class category within preformed set of traffic signs. The two main modules of our proposed method will be described briefly in the following subsections respectively.
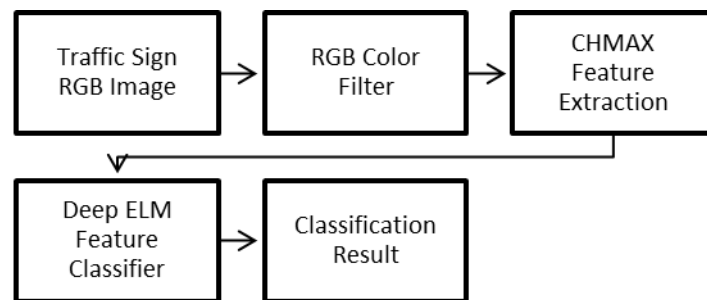

Figure 1. Flowchart of the Proposed Network Framework

2.2. Color HMAX (CHMAX) Based Traffic Sign Feature Extractor

The originally proposed HMAX model [14] for the features extraction was in grey scale. In many later studies, many researchers proposed different variants of HMAX model to include color cues for the features extraction from color images and improvement in accuracy performance was reported. In our earlier work, we proposed a novel feature processing framework for the joint processing of the shape and color features based on the fusion of the HMAX model and the color opponent channels [17] and the CHMAX features output were fed to the support vector machine classifier to classify the RGB image to its respective class category. Recently, extreme learning machine was developed by Guang-Bin-Huang *et al.* [12] as an alternative to support vector machine SVM, and it was proved to be efficient enough to compete with SVM classifiers in terms of speed, accuracy and reliability. For a wide range of complex patterns, deep neural networks are favorable as they are more efficient in representing complex patterns with much variance due to their good generalization capability. Therefore we developed DELM deep extreme learning machine based feature classifier to

further improve the recognition and classification of our earlier work [17] because it is comparatively less computation extensive and trains comparatively faster.
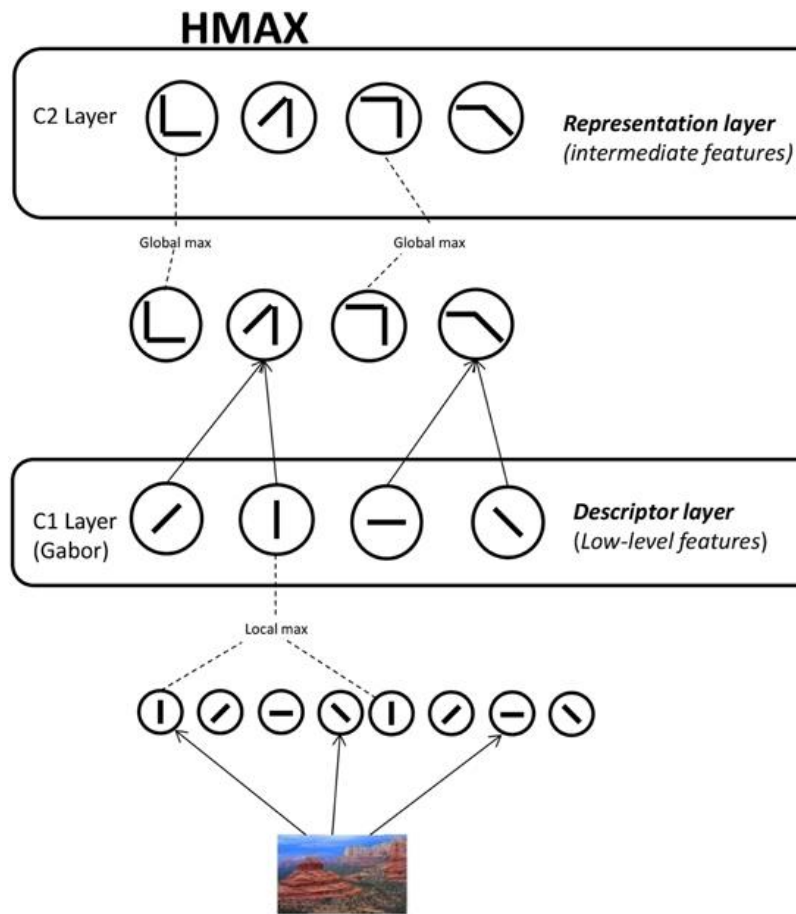

Figure 2. Architecture of the Famous HMAX Model

In this paper, we proposed a RGB opponent color channel color descriptor and integrate it with the famous computer vision model known as HMAX shown in Figure 2. The proposed feature descriptor comprises of two stages. Traffic sign texture and color features are extracted in stage I, by simulating the neuronal function of the human vision system. Whereas the traffic sign recognition and classification occurs in stage II, also by simulating the neuronal function of the human learning intelligence system. Three equi-luminant chromatic axes are used in our proposed color descriptor based on the neurological evidences namely, Red-Green, Yellow-Blue and Black-White channel. The Black-White channel is the combination of red green and blue color channels together, while yellow is the combination of red and green (see Figure 3). The red, green and blue color components were combined into 3 pairs of color channels using a weight matrix Ω_{3c} represented as follows:

$$\Omega_{RGYBWB} = \begin{pmatrix} \pm 1/\sqrt{2} & \pm 1/\sqrt{6} & \pm 1/\sqrt{3} \\ \mp 1/\sqrt{2} & \pm 1/\sqrt{6} & \pm 1/\sqrt{3} \\ 0 & \mp 2/\sqrt{6} & \pm 1/\sqrt{3} \end{pmatrix}$$

where individual color component weights $\omega = (\omega_R; \omega_G; \omega_B)$ for each of the three color components are stored as a column vector as shown. The first column in the above weight matrix shows the individual color component weight values for Red-Green channel ω_{RG},

while the second column represents the individual color component values for Yellow-Blue ω_{YB} and the third column represents weight vectors for White-Black ω_{WB} channel. In the weight vector shown above, $\omega_{YB} = (\pm 1/\sqrt{6}, \pm 1/\sqrt{6}, \mp 2/\sqrt{6})$ such that the "+"

(respective "-") sign indicates an excitatory yellow center (made from the combination of red and green) and inhibitory blue surround component. All the parameters used in the implementation of the single opponent and double opponent color channels are constrained by the neuroscience data.
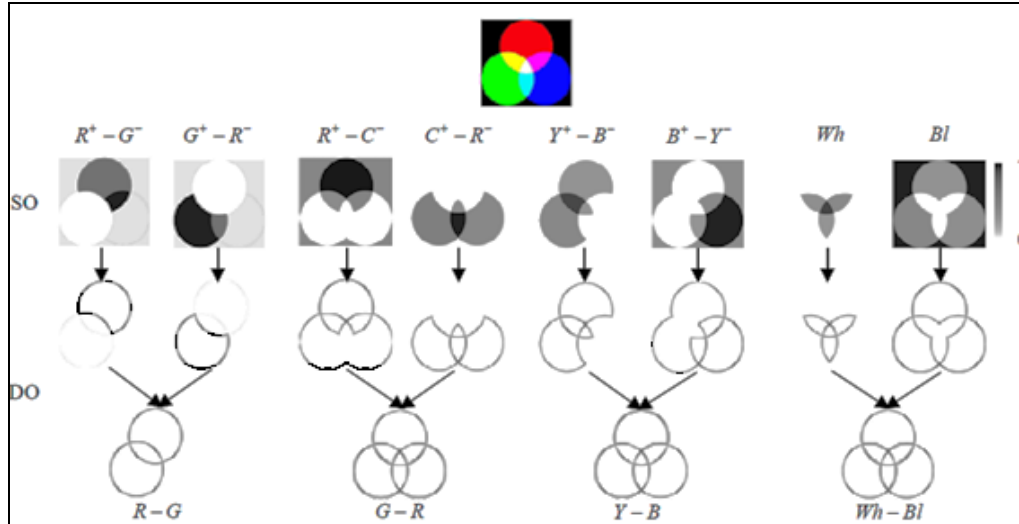

Figure 3. Composition of Opponent Color Channels from a RGB Image

Following the creation of three opponent color channels with the proposed color descriptor, the traffic sign texture features are extracted from the traffic sign RGB image using HMAX based traffic sign pattern feature descriptor. The famous computer vision model known as HMAX, consists of 4 layers in a feed-forward hierarchy namely, S1, C1, S2 and C2 layers. These layers imitate the hierarchy of layers in the primary visual cortex. These layer follow an alternating sequence of convolution operation in one layer and the MAX pooling operation in another layer. Feature maps are constructed using the convolution operation in each convolution layer. Whereas robustness to geometric transformations like changes in scale, position and orientation of the traffic sign is achieved by the MAX pooling operation. We implemented a dictionary of thousand features for each image at sixteen scales and four orientations. The output of CHMAX is then fed to the deep ELM based traffic sign classifier module as shown in Figure 1, for the classification of the traffic sign RGB image to their respective class category. Further details regarding the feature extraction module are described in detail in [17]. In this paper, we will discuss DELM extensively in the following subsection.

2.3. Deep ELM (DELM) Based Traffic Sign Classifier

Figure 4. D-Modules Deep ELM (DELM) Network

In our DELM model, we construct a deep ELM network feature classifier by stacking standard ELM modules (three layer feed-forward model input layer, one hidden layer and output layer) in cascade as shown in Fig 4, each following ELM module receives its input from the preceding module. Therefore a D-module deep ELM (DELM) is implemented adding 'D' identical ELM modules in cascade.

As described earlier, traffic sign features extracted by the CHMAX module are then fed to the DELM classifier, which acts as an input vector $X \in \mathbb{R}^{N \times 1}$ to the deep ELM network. The input weight matrix connecting the input layer with the hidden layer is defined by $W_{\rho 1} X \in \mathbb{R}^{M \times 1}$. The lth hidden layer response is acquired by the sigmoid function as follows:

$$f[H_{1,l}] = \frac{1}{1 + exp(-H_{1,l})}. \tag{1}$$

An approximation of the input, $X \in \mathbb{R}^{N \times 1}$ is then produced at the output by multiplying the hidden layer response with the output weight matrix W_{O1}. Auto-encoding is performed by using K training vectors to train the output weights, we define a matrix $G \in \mathbb{R}^{M \times K}$ where each column holds the hidden layer response $f[H_1]$ at 1 training point. Then we solve for $W_{O1} \in \mathbb{R}^{N \times M}$ to minimize the mean square error using the training data $Y \in \mathbb{R}^{N \times 1}$ where each column holds training vectors, and the set of original training images, Y, Similar to supervised training of an ELM carried out elsewhere [18], we find the solution of the following set of NM linear equations in NM unknown variables comprised from the elements of W_{O1},

$$Y_{predicted} := W_{O1}G \tag{2}$$
$$YA^T = W_{O1}(AA^T + cI), \tag{3}$$

where c is the regularization parameter that can be optimized as a hyper parameter, and I is an identity matrix of order (M).

The auto-encoded vector form $\overline{X_1}$ of the input instance X is calculated with the help of the trained weight matrix W_{O1} by $\overline{X_1} = W_{O1}f[H_1]$.

Then the second module of the deep network is constructed using weight matrix $W_{p2} \in \mathbb{R}^{Q \times N}$, $\hat{X}$ is connected to the 2nd hidden layer, whose input is defined by $H_2 \in \mathbb{R}^{Q \times 1}$, where Q the size of second hidden layer is not necessarily the same as the size of first hidden layer, M. Then the second module output weights, W_{O2}, are also trained in the same way as of the first module to produce a new auto-encoded response, $\overline{X_2} = W_{O2}f[H_2]$.

The entire process can be repeated numerous times, to form a D-module deep ELM network.

The procedure required for the training of weights of the deep ELM network is explained above. After the training process is completed, the 2 weight matrices W_{O1}, W_{O2} are combined to form a single weight matrix $W_{h12} \in \mathbb{R}^{Q \times M}$, connecting the 1st two hidden layers as,

$$W_{h12} = W_{p2}W_{O1}, \tag{4}$$

Similarly, this step can be repeated sequentially for subsequent modules, since both the input and output layers of the ELM network are linear.

2.4. Parameter Selection of the Deep ELM Network

The input and output dimension (N) for the deep ELM network was 1000 features for each image. We chose number of non-linear neurons (M) in a hidden layer as 12,800 and number of ELM modules in cascade (D) equals 3. (see [17, 19] for detailed explanation for the selection of values for optimum results)
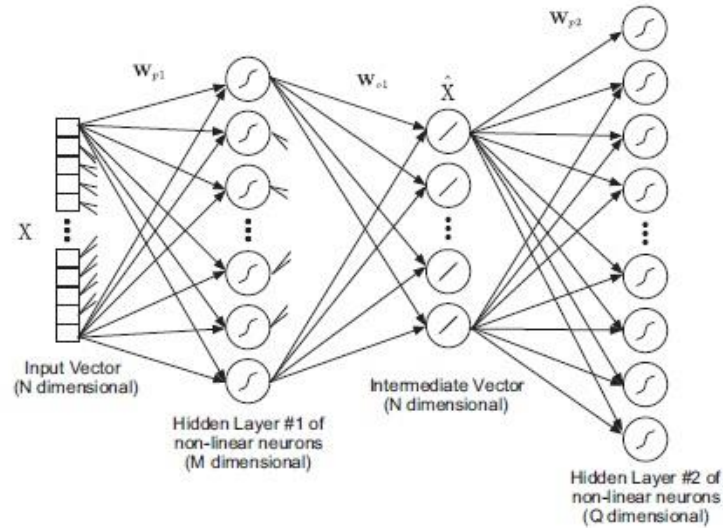
The approach described above is illustrated in Figure 5.

Figure 5. Deep ELM Network Architecture

3. Experiment

3.1. Dataset

Figure 6. German Traffic Sign Images

To evaluate the recognition accuracy of the developed algorithm, we used the German Traffic Signal Recognition Benchmark dataset. GTSRB consist of 50000 annotations from 43 classes of traffic signs. The resolution of each image is 1360 x 800 pixels.

The accuracy of the developed network was calculated based on the statistical criteria as follows:

$$classification\ accuracy = \frac{number\ of\ correct\ classifications}{total\ number\ of\ testing\ images}$$

The recognition performance of the proposed method is calculated using ten cross-validation method. Entire data set is divided into ten segments. Each segment is used for testing turn by turn, the remaining nine segments were used for training in each turn. Finally the ten results were averaged together to calculate the average performance of the proposed method over the entire data set.

4. Results

The traffic sign features are extracted using biologically inspired proposed algorithm and then recognized successfully and classified automatically to the respective class category using deep (multiple hidden layers) extreme layer machine. The automatic recognition and classification accuracy of the proposed improved model is calculated and then compared with the earlier proposed model [17,20-21] (See Table 1& 2).

We have shown the recognition accuracy of the developed network with different number of hidden nodes in Table 1. It is evident from Table 2 that the use of deep extreme learning machine as a classifier achieved better recognition accuracy as compared to the other existing competitive algorithms. The developed algorithm is found to be capable of providing robust recognition and accurate classification. Hence, we conclude that the

developed algorithm performs invariant recognition and classification of traffic signs efficiently.

Table 1. D-ELM Recognition Accuracy with Different Number of Hidden Nodes

Hidden Nodes	2000	4100	6200	8400	10600	12800
Rec. Accuracy	98.70%	99.10%	99.20%	99.30%	99.35%	99.40%

Table 2. A Comparison of Recognition Accuracy of our Study with other Existing Computer Vision Algorithms

Average Accuracy	Method			
	Co-occurrence Method*	Color HMAX with SVM Classifier*	Color HMAX with conventional ELM	Color HMAX with multiple hidden layer (deep ELM)
	87.77%	98%	98.2%	99.4

*Recognition accuracy is taken from quoted references [17, 20-21].

Earlier, we proposed a model for feature extraction and classification from color image using support vector machine as a classifier [17]. In order to further improve the classification accuracy of the earlier proposed algorithm, in this paper we proposed a deep ELM based classifier model and integrated this classifier into the earlier proposed model by replacing SVM classifier. We have shown a comparison of the proposed algorithm with other implementations ([17, 20-21]) in Table 2. It can be seen clearly, that the biologically plausible traffic sign analysis method is much efficient and accurate than other existing computer vision models. It is evident from Table 2, that extreme learning machine is a good alternative to Support Vector Machine classifier, and is biologically plausible as well. Furthermore, we noticed that a multi-hidden layer variant of ELM further improves the recognition and classification performance. Hence we believe that use of extreme learning machine as a classifier to construct a biologically inspired network to imitate human vision and human intelligence, is a great leap forward to achieve human-like efficient intelligence.

5. Conclusion

We have proposed a biologically inspired model of joint processing for extraction of traffic sign color and shape information based on known properties of the primate visual cortex, capable of recognizing and classifying traffic signs invariantly from color image. We have tested the algorithm to recognize traffic signs from the color images in the GTSRB, and shown that the proposed descriptor perform better than the existing algorithms. The proposed model was shown to yield higher accuracy for automated recognition and classification of traffic signs. Yet, most of the errors in recognition mainly occurred sue to motion blur implies that the performance may further be improved by adding some layers that may detect features invariant to blur effect. Overall the relative success of the proposed biologically inspired approach suggests that neuroscience may contribute new ideas and superior algorithms for automatic recognition and classification applications using color images.

Acknowledgement

References.

[1] Timofte, R., Prisacariu, V.A., Van Gool, L.J., & Reid, I., 2011, In C.H.Chen (Ed.), Combining traffic sign detection with 3d tracking towards better driver assistance, World Scientific Publishing

[2] Levinson, J., Askeland, J., Becker, J., Dolson, J., Held, D., Kammel, S., Kolter, J.Z., Langer, D., Pink, O., Pratt, V., Sokolsku, M., Stanek, G., Stavens, D., Teichman, A., Werling, M., & Thrun, S. (2011), Towards fully autonomous driving: System and algorithms, In Proceedings of 2011 IEEE intelligenct vehicles symposium (pp. 163-168)

[3] Mogelmose, A., Trivedi, M.M., & Moeslund, T.B. (2012), Vision-based traffic sign detection and analysis for intelligent driver assistance systems: perspective and survey, IEEE Transactions on Intelligent Transportation Systems, 13, 1484-1497.

[4] Maldonado-Bascon, S., Lafuente-Arroyo, S., Gil-Jimenez, P., Gomez-Moreno, H., & Lopez-Ferreras, F., (2007), Road-sign detection and recognition based on support vector machines, IEEE Transactions on Intelligent Transportation Systems, 8, 264-278.

[5] delaEscalera, A., Moreno, L.E., Salichs, M.A., & Armingol, J.M., (1997), Road traffic sign detection and classification, IEEE Transactions on Industrial Electronics, 44, 848-859.

[6] Khan, J.F., Bhuiyan, S.M.A., & Adhami, R.R., (2011), Image segmentation and shape analysis for road sign detection, IEEE Transactions on Intelligent Transportation Systems, 12, 83-96.

[7] Gavrila, D.,(1999), Traffic sign recognition revisited, 21, DAGM-Sympasium (pp.86-93), London, UK, Springer-Verlag

[8] Salti, S., Patrelli, A., Tombari, F., Fioraio, N.,& Di Stefano, L., (2015), Traffic sign detection via interest region extraction, Pattern Recognition, 48, 1039-1049.

[9] Lillo-Castellano, J.M., Mora-Jimenez, I., Figuera-Pozuelo, C.,& Rojo-Alvarez, J.L., (2015), Traffic sign segmentation and classification using statistical learning methods, Neurocomputing, 153, 286-299.

[10] Greenhalgh, J.,& Mirmehdi, M., (2012), Real-time detection and recognition of road traffic signs, IEEE Transactions on Intelligent Transportation Systems, 13, 1498-1506.

[11] Schmidt, Wouter, F., Martin A. Kraaijveld, and Robert PW Duin, (1992), Feed-forward neural networks with random weights, Pattern Recognition, 1992, Vol. II. Conference B: Pattern Recognition Methodology and Systems, Proceedings, 11th IAPR International Conference on IEEE.

[12] Guang-Bin-Huang, Qin-Yu Zhu and Chee-Kheong Siew, (2004), Extreme learning machine: a new learning scheme of feed-forward neural networks, Neural Networks, 2004, Proceedings IEEE International Joint Conference on Vol. 2.

[13] J. Tang, C. Deng, G.-B. Huang and B. Zhao, (2014), Compressed Domain-Ship Detection on Spaceborne Optical Image Using Deep Neural Network and Extreme Learning Machine. IEEE Transactions on Geoscience and Remote Sensing.

[14] T. Serre, L. Wolf, S.M. Bileschi, M. Reisenhuber, T. Poggio, (2007), Robust object recognition with cortex-like mechanism, IEEE Transactions on Pattern Analysis and Machine Intelligence, 29, 411-426.

[15] Conway, Bevil R., (2010), Advances in color science: from retina to behavior. The Journal of Neuroscience 30.45: 14955-14963.

[16] Stallkamp, J., Schlipsing, M., Salmen, J., & Igel, C.,(2011), The german traffic sign recognition benchmark: a multiclass classification competition, In proceedings of 2011 International Joint Conference on Neural Networks (IJCNN), pp. 1453-1460.

[17] Babar Khan, Fang Han, Zhijie Wang, Rana J Masood, (2016), Bio-Inspired Approach to Invariant Recognition and Classification of Fabric Weave Patterns and Yarn Color, Assembly Automation, Vol 36, Issue 2, pp 152-158.

[18] G.-B. Huang, D.H. Wang, Y. Lan, (2011), Extreme learning machines: A survey, Int. J. Mach. Learn. Cybern. 2, 107–122.

[19] M.D. Tissera, M.D. McDonnell, (2016), Deep extreme learning machines: supervised autoencoding architecture for classification, Journal of Neurocomputing, vol. 174, 42-49.

[20] P.F. Li, J. Wang, H.H. Zhang and J.F. Jing, (2012), Automatic Woven Fabric Classification Based on Support Vector Machine, International Conference on Automation Control and Artificial Intelligence, Xiamen, China, 581-585.

[21] B.S. Yassine and N. Salem, (2014), Automatic Recognition of Woven Fabrics Based on Texture and Using SVM, Signal Image and Video Processing, Vol. 11, 429-434.