



Cognition based recognition of partially occluded traffic signs

A. Mannan^{a,b,*}, K. Javed^b, A. ur Rehman^b, H.A. Babri^b, S.K. Noon^a

a. *Department of Electrical Engineering, NFC Institute of Engineering and Technology, Multan, 59060, Pakistan.*

b. *Department of Electrical Engineering, University of Engineering and Technology, Lahore, Pakistan.*

Received 9 January 2020; received in revised form 15 December 2021; accepted 14 February 2022

KEYWORDS

Traffic sign;
 Cognition;
 Occlusion;
 Parameter estimation;
 Discrete cosine transform;
 Dimensionality reduction;
 Feature extraction,
 Convolutional neural networks.

Abstract. Computer vision-based traffic sign detection and recognition is an active field of research, but the task becomes challenging when the sign of interest is partially occluded by nearby objects like a tree, pole, or vehicle. Another difficulty posed especially in the developing countries is the lost colors problem that arises from aging and poor maintenance. This work presents an automatic technique that focuses on visible parts only and suppresses occluded portions. Features are collected using a convolutional neural network inspired invariant feature extraction technique augmented with feature interaction-based dimensionality reduction. Further, with the use of dynamic parameter estimation, an adaptive system for continuous learning is also proposed. Since the effect of partial occlusion has not been thoroughly studied, there is no benchmark database available for this purpose. We have prepared two datasets by combining originally and synthetically occluded images taken from field surveys and from famous GTSRB database. Experiments revealed that our technique outperformed state-of-the-art recognition methods previously used for visible and occluded signs by obtaining 0.81 precision and 0.79 recall values, on average. The proposed method also exhibits a remarkably low error rate as the amount of occlusion increases.

© 2022 Sharif University of Technology. All rights reserved.

1. Introduction

Traffic signs are used worldwide to guide road users with an objective to avoid accidents [1]. Mainly, there are two types of traffic signs:

1. American with white background and black foreground;
2. European having either red rim or filled with blue color [2–5].

In this work, we focus on European traffic signs only. Many computer vision-based methods used to automatically detect and recognize European traffic signs

have been reported in literature [1,6–10]. Detection serves the purpose of segmenting a traffic sign in a real-world scene, whereas recognition deals with reading its contents. Many automatic traffic sign detection and recognition systems can effectively detect and recognize visible signs in many parts of the world [11]. For detection, various colors, saliency, and statistical distribution of pixel-based methods have been reported in the literature [1,12,13]. The next subproblem, i.e., recognition, has been tackled by researchers by using different techniques to extract invariant features from traffic sign shapes. Some famous methods are Histogram of Oriented Gradients (HOG) [14], Local Binary Patterns (LBP) [15], their combinations [1], transform based methods [1] (e.g., Fourier, discrete wavelet and discrete cosine etc.), and higher order spectra [16]. Feature extraction can be applied to an image as a whole or with the help of a rectangular

*. *Corresponding author. Tel.: 00923216301074
 E-mail address: mannan@nfciet.edu.pk (A. Mannan)*

or polar grid [2]. Defined as reducing the size of feature vector by removing irrelevant and redundant features, dimensionality reduction [17] has proved to be an effective method. Dimensionality reduction not only enhances computational speed but also increases the recognition accuracy [18]. A multiclass classifier is then employed to assign traffic signs their correct classes [19]. An alternate approach to recognizing contents of a traffic sign is using deep neural networks [20,21] which embed both feature learning and classification stages. However, internal architecture of such networks is complex and requires learning of thousands (or even millions) of parameters. On small datasets containing only a few thousand images, these networks are prone to over-fitting [22]. Pre-trained models of famous Convolutional Neural Networks (CNN) like Alexnet or Googlenet [23,24] can, however, be customized to fit for a small dataset [25].

In some cases, a traffic sign may be occluded by other nearby objects [26,27]. This occlusion can be caused by a tree, vehicle, pole or be deliberate, e.g., by pasting a sticker or writing something on the sign board. Detecting and recognizing an occluded traffic sign is challenging because some portion of the object of interest is not visible. For partially occluded traffic sign images, extracting meaningful features is a bit tricky because we have to take care of the occluded portion too. We have found a number of attempts at detection of occluded traffic signs in the literature including the novel pre-processing layer-based solution proposed by the current author [28], but there is indeed quite a limited scope of relevant research on the recognition of their contents.

We have identified the four following categories of occlusions commonly found for traffic signs:

- Natural,
- Incidental,
- Deliberate,
- Degradation.

Natural occlusion is caused by a nearby tree branch or leaves; incidental is the occlusion by chance due to the presence of a vehicle or pole. The third category of occlusion occurs when humans deliberately suppress the contents of a traffic sign by putting an advertisement sticker on it. In certain parts of the world where the sunshine is bright and/or due to lack of adequate maintenance, traffic signs are frequently found degraded (i.e., Category 4). Examples of each of the four types of occlusions are shown in Figure 1.

We, in this work, address the problem of recognizing contents of occluded and degraded traffic signs by taking inspiration from human cognition. Instead of working in Red Green Blue (RGB) domain, we obtain a boosted image with the help of linear mapping and

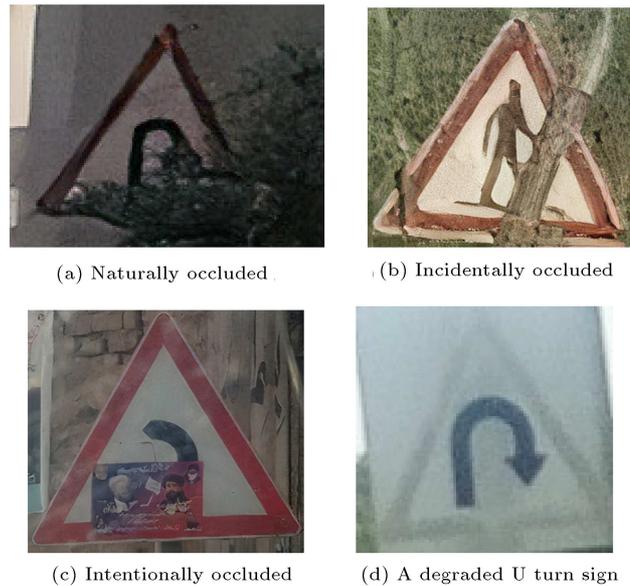


Figure 1. Different categories of occlusions found in real-world traffic signs of occluded traffic signs only.

location aware edge filtering. In this image, portions related to the occlusions are suppressed and the visible parts are highlighted so that our CNN inspired multiscale spectral feature extraction method can concentrate on visible portions only. A dimensionality reduction technique using feature interaction has also been employed to reduce the size of data which results in lower computational cost and higher recognition accuracy. Finally, a multiclass Support Vector Machine (SVM) classifier is used to make predictions. In order to address the subproblem of degradation, a dynamic parameters estimation technique with feedback is also proposed.

The remainder of this paper is organized as follows: Section 2 describes some related work especially pertaining to occluded and degraded traffic sign's detection and recognition. Section 3 describes our proposed technique to recognize occluded and degraded traffic signs. Section 4 explains experimental setup. Section 5 describes the results obtained using different settings. Section 6 discusses experimental results of our proposed method, and finally Section 7 describes conclusion of the work.

2. Related work

A detailed review of standard and visible traffic sign detection and recognition techniques can be found in [1,29]. We, in this section, limit our discussion to detection and recognition.

Though automatic traffic sign detection and recognition is an active field of research, dealing with occluded and degraded signs has been addressed the least. We found only few attempts to handle this

problem in literature. Occlusion maps based technique in [27] uses decomposition of SVM score to detect possible traffic signs in a given real-world scene, but the work is limited to the detection and super class identification, i.e., circle, triangle, etc. Occlusion maps were originally proposed by Wang et al. [30] to detect occluded objects in images; they used a combination of HOG and LBP to detect occluded objects of interest. Rehman et al. [31] used a scheme to automatically identify occluded portions during training process by using a predefined heuristic. Then, the features were collected of visible portions only called discriminant patches (d-patches). A disadvantage of this approach is that the performance degrades severely with increasing amount of occlusion on test data. A similar work for detecting traffic signs in real-world scenes was presented in [15]. They proposed the use of a large number of color channels from various color spaces and applied LBP as feature extraction technique. Finally, integral features were used to generate a large feature vector. Floros et al. [32] developed a scheme to detect degraded traffic signs in natural scenes using RGB thresholding and heuristic based search. For experimentation, they collected a small dataset of around three hundred images divided into only five classes from two different places in Greece. Li and Ma [33] used a faster version of probabilistic neural networks to recognize occluded and degraded traffic signs on a synthetic dataset of few hundred images, but it was prone to overfitting due to small training and test datasets. A generative model was trained and used by Ishida et al. [34] to identify degraded traffic signs taken from a synthetically generated dataset of standard and degraded traffic signs. All methods mentioned above are mainly used to detect partially occluded and/or degraded traffic signs in real-world scenes leaving recognition of contents as a future work.

To recognize contents of partially occluded traffic signs, some relatively older systems designed for standard and visible signs [26,35] also respond well. Fleyeh and Davami [26] made use of eigen space and considered only top few eigen vectors as features; these were found effective for some partial occlusions too. De La Escalera et al. [35] used deformable models based on energy of color and gradient to detect traffic sign in natural scenes, but they used a very small dataset for evaluation.

Keeping in view partial occlusions, we have found a number of techniques to identify objects in general, but there are only a few methods pertaining to traffic signs' detection and recognition.

3. Our proposed technique

In this section, our proposed method for recognition of partially occluded traffic signs is described at length.

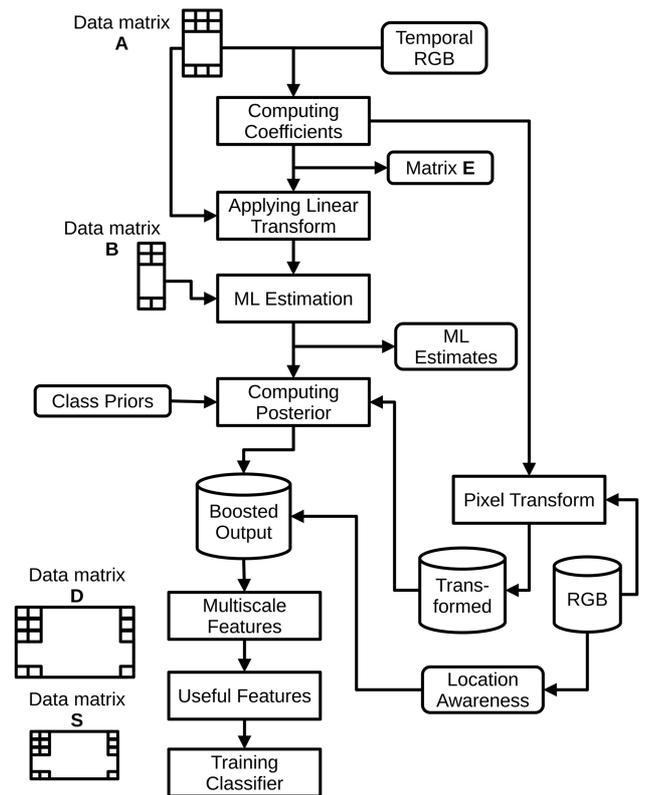


Figure 2. Block diagram of the proposed system.

Block diagram of the proposed system is shown in Figure 2 and a brief description of different blocks is given in the following sections.

3.1. Linear transformation

The objective of linear transformation is to map RGB input data to another color space where:

- Pixels corresponding to different colors are mapped far from each other;
- Pixels corresponding to the same color are mapped very close to each other.

As shown in the top left corner of Figure 2, a data matrix **A** is created by putting RGB data corresponding to red and blue colors as its rows. The data in each row is then transformed to an output matrix **B** according to the linear transformation mentioned in Eq. (1):

$$\mathbf{b}(j) = T_{lin}[\mathbf{a}(j)] = \mathbf{a}(j) \times \mathbf{E}. \quad (1)$$

At any arbitrary row j in the data matrix **A**, $\mathbf{a}(j) = [a_1(j) \ a_2(j) \ a_3(j)]$ is a tuple containing R, G, and B color channels, whereas $\mathbf{b}(j) = [b_1(j) \ b_2(j)]$ is the dataset at the same row in the transformed matrix **B**. $\mathbf{a}(j)$ and a 1×3 vector is multiplied by the transformation matrix **E** of size 3×2 to give $\mathbf{b}(j)$ of size 1×2 .

3.1.1. Computing the coefficient of matrix \mathbf{E}

To compute the coefficients of matrix \mathbf{E} , the RGB pixel data contained in matrix \mathbf{A} is mapped on the Eigen space and the two vectors pertaining to the largest two Eigen values are retained [36,37]. The process includes the following steps:

- Covariance matrix $\Sigma^{\mathbf{A}}$ (of size 3×3) was computed for the data matrix \mathbf{A} (of size $n \times 3$), where n is the number of rows in matrix \mathbf{A} ;
- Three Eigen values λ_1, λ_2 , and λ_3 were computed for the covariance matrix $\Sigma^{\mathbf{A}}$ such that $\lambda_1 > \lambda_2 > \lambda_3$;
- Two Eigen vectors \mathbf{w}_1 and \mathbf{w}_2 corresponding to the two largest Eigen values λ_1 and λ_2 were computed and put as the columns of the matrix \mathbf{E} i.e., $\mathbf{E} = [\mathbf{w}_1 \ \mathbf{w}_2]$. This gives the matrix \mathbf{E} of size 3×2 .

This linear transformation attempts to maximize separation among colors pertaining to the traffic signs and those taken by the occlusions and the background in addition to the advantage of reducing input data dimensions from three to two [38].

3.2. Dynamic parameter estimation

As a result of applying Eq. (1) on each row of 3-dimensional data matrix \mathbf{A} , a 2-dimensional transformed matrix \mathbf{B} is obtained. The next step is to estimate parameters of the two color classes (k), assuming each of them to be normally distributed, i.e.:

$$P(\mathbf{b}|c_k) \sim \mathcal{N}(\mathbf{b}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \quad k = 1, 2. \quad (2)$$

Maximum likelihood estimates [39] of multivariate mean $\boldsymbol{\mu}_k$ and covariance $\boldsymbol{\Sigma}_k$ of training data are given in Eqs. (3) and (4). It is assumed that there are R training images having P pixels of the color k per training image. As mentioned earlier, $\mathbf{b}(j)$ is the two-dimensional data set present at location j in data matrix \mathbf{B} . To be more precise, $\mathbf{b}_{r,p,k}(j)$ is a tuple of two-dimensional data set corresponding to the p th pixel in the r th training image pertaining to class k present at any arbitrary location j in data matrix \mathbf{B} . Since index j is obvious, we will drop it for subsequent discussions. The superscript t in Eqs. (3) and (4) shows the time index used to update these estimates. Since the system is capable to adapt under any situation, the RGB color data is collected at run time and is appended at the end of matrix \mathbf{A} (refer to Figure 2) if it passes a statistical test. Details of updating maximum likelihood parameters at run time are discussed in Section 4.

$$\boldsymbol{\mu}_k^t = \frac{\sum_{r=1}^R \sum_{p=1}^P \mathbf{b}_{r,p,k}^t}{R \times P}, \quad (3)$$

$$\boldsymbol{\Sigma}_k^t = \frac{\sum_{r=1}^R \sum_{p=1}^P (\mathbf{b}_{r,p,k}^t - \boldsymbol{\mu}_k^t) (\mathbf{b}_{r,p,k}^t - \boldsymbol{\mu}_k^t)^T}{R \times P}. \quad (4)$$

3.3. Boosted image

Given that an image \mathbf{G} on the transformed color space is obtained by applying Eq. (1) to RGB image, the likelihood function mentioned in Eq. (2) is used to know how likely each pixel in the transformed domain image \mathbf{G} belongs to class c_k . With this information available, posterior probability is computed using Bayes rule mentioned in Eq. (5). $P(\mathbf{g}(x, y)|c_k)$ is the likelihood that pixel $\mathbf{g}(x, y)$ belongs to class c_k , $P(\mathbf{g}(x, y))$ is evidence, and $P(c_k)$ is the prior probability. The prior depends on the proportional representation of samples from each class and is computed using Eq. (6). Here, N_k is the number of instances in the training data belonging to class c_k and N_T is the total number of training instances.

$$P(c_k | \mathbf{g}(x, y)) = \frac{P(\mathbf{g}(x, y)|c_k)P(c_k)}{P(\mathbf{g}(x, y))}, \quad (5)$$

$$P(c_k) = \frac{N_k}{N_T}. \quad (6)$$

Posterior probability $P(c_k | \mathbf{g}(x, y))$ corresponds to a grayscale image in which pixels corresponding to the colors of interest, i.e., red and blue, are intended to have significantly higher gray values than other pixels. Red and blue colors are chosen because the European traffic signs we are interested in appear either with a red outer rim or are filled with blue color [40]. In order to highlight the contents that are either black foreground on white background or white foreground on blue background, location aware edge filtering $edge_1^h[\mathbf{g}(x, y)]$ is employed. Vector \mathbf{l} is a set of spatial locations to perform filtering, while h is the minimum edge strength to be retained. Values of \mathbf{l} and h were empirically computed from the training data. By using this technique, high-contrast edges present only inside the traffic sign rim, i.e., middle of the shape, are highlighted, whereas all other edges are suppressed. As a result, the pixels corresponding to the contents of the sign take higher gray values than the rest. It is to be noted that any occlusions present in the middle of the shape having weak edges are also suppressed.

The operation $edge_1^h[\mathbf{g}(x, y)]$ is further explained with the help of a graphical example shown in Figure 3 for different values of h and I . The subscript “Low” indicates that even very weak edges were captured, whereas the subscript “High” indicates that only the very strong edges were retained. “Medium” is the edge strength in between the former two options. It can be clearly seen that the best output of the location-aware edge filtering can be ensured if even very strong edges in the background are suppressed and the medium edge strength in the foreground is retained (refer to Figure 3(e)).

Finally, the sum of data corresponding to the posterior, i.e., $P(c_k | \mathbf{g}(x, y))$, and location aware edge



Figure 3. Graphical description of the proposed location-aware edge filtering by varying parameters h and I . The filtered image obtained using I_{High} and h_{Medium} produced the best results.

filtering $edge_1^h[\mathbf{g}(x, y)]$ is linearly transformed to form the gray values of the desired boosted image \mathbf{I} , mathematically:

$$i(x, y) = 255 \times P(c_k | \mathbf{g}(x, y)) + edge_1^h[\mathbf{g}(x, y)], \quad (7)$$

where $i(x, y)$ is a scalar number equal to the gray value of Image \mathbf{I} at location (x, y) .

In the boosted image \mathbf{I} , the pixels representing the traffic sign take higher gray levels than the rest. An example is shown in Figure 4 where a sample occluded image shown in Figure 4(a) was processed to obtain the boosted image \mathbf{I} shown in Figure 4(b). Further, Figure 4(c) shows the surface plot of the boosted image showing that pixels corresponding to the traffic sign take on higher gray values and the effect of occlusion and background is suppressed.

3.4. Multiscale spectral feature extraction

The ability of Discrete Cosine Transform (DCT) for feature extraction in images has already been ex-

plored [2,41–43]. Ayyalasomayajula et al. [42] proposed DCT phase (i.e., sign, either positive or negative) based descriptor to recognize partially occluded real-world images. To use for partially occluded and degraded traffic sign recognition, we extend their method as follows:

- In order to capture detailed features, we, unlike the study in [42], apply feature extraction method on multiple spatial scales to the boosted image \mathbf{I} instead of the original RGB image. In Subsection 3.3, we mentioned that in the boosted image, the effect of occlusion has already been partially suppressed, which makes the task of feature extraction technique easier;
- Since the resulting descriptor is high dimensional, Ayyalasomayajula et al. [42] used empirically determined thresholds to select important features. We, on the other hand, use a more robust technique, namely feature interaction based dimensionality re-

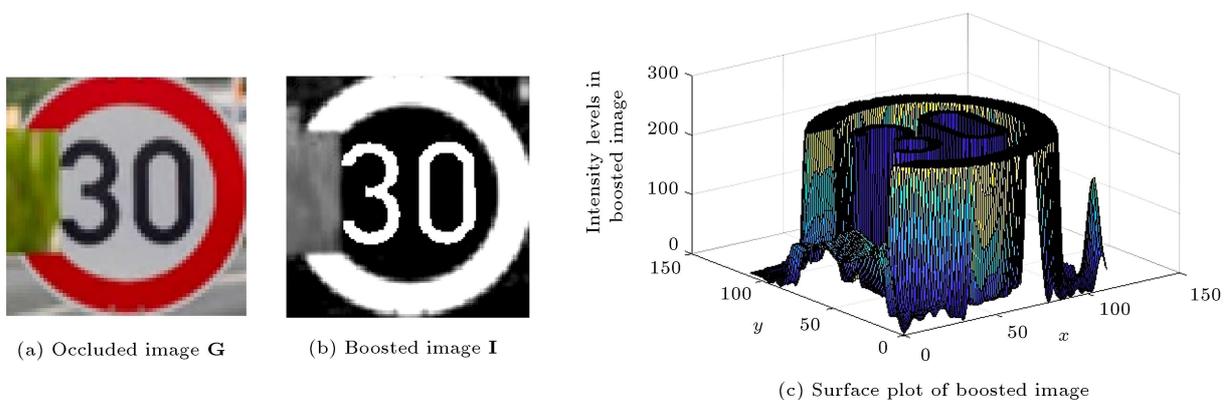


Figure 4. A sample occluded image and its boosted version. Pixels corresponding to the background and occlusions take significantly lower gray values than the pixels corresponding to the traffic sign.

duction, to select the most relevant and the least redundant information.

Given a boosted spatial domain image \mathbf{I} , let us represent the DCT of a pixel $i(x, y)$ as $i(u, v)$ by using indices (x, y) for spatial and (u, v) for frequency domain. Mathematical expression to compute the transform of an image of size $X \times Y$ is given in Eq. (8) where α_u and α_v are simple functions of indices u and v [44]. The real valued pixel in transformed domain, i.e., $i(u, v)$, can be written in terms of magnitude component which is its absolute value $i_{abs}(u, v)$ and the phase term that is the sign either positive or negative $i_{ph}(u, v)$, as stated in Eq. (9):

$$i(u, v) = T_{DCT}[i(x, y)]$$

$$= \alpha_u \alpha_v \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} i(x, y) \cos \frac{\pi(2x+1)u}{2X} \cos \frac{\pi(2y+1)v}{2Y}, \quad (8)$$

$$0 \leq u \leq X - 1 \quad 0 \leq v \leq Y - 1,$$

$$i(u, v) = i_{abs}(u, v) \times i_{ph}(u, v), \quad (9)$$

where:

$$i_{abs}(u, v) = \|i(u, v)\|, \quad (10)$$

and:

$$i_{ph}(u, v) = \frac{i(u, v)}{i_{abs}(u, v)}, \quad (11)$$

and hence:

$$\mathbf{I} = \mathbf{I}_{abs} \circ \mathbf{I}_{ph}, \quad (12)$$

where “o” refers to the Hadamard product operation [45] where element wise multiplication is performed between the two matrices \mathbf{I}_{abs} (containing only the magnitude) and \mathbf{I}_{ph} (containing only the phase). Figure 5 shows a sample input image containing a traffic sign and the images reconstructed from the magnitude and phase of its DCT. It can be seen that the phase reconstructed image [46] contains more

visual information than magnitude reconstructed image. This renders the DCT transformed phase image more suitable for the purpose of extracting invariant features [42].

In order to obtain an occlusion invariant descriptor, we propose a feature extraction strategy inspired by famous CNN [12]. Block diagram of the feature extraction technique is shown in Figure 6. Each input image in the training set is passed through two average pooling layers where average operation is performed on a box of 2×2 pixels which in turn reduces the size of the image to one half for the subsequent layer (i.e., filter size = 2×2 and stride = $2f$). DCT phase features are computed on images tapped at different intermediate points (i.e., scales) and the features are finally concatenated. It should be noted that in each layer, feature extraction operation is performed on blocks of size 8×8 which was empirically found to be the best choice.

Mathematical representation of this operation on a sample image \mathbf{I}_i taken from the training dataset is shown in Eq. (13):

$$\mathbf{d}_i = cat\{T_{DCT_{ph}}^{64 \times 64}[\mathbf{I}_i] \quad T_{DCT_{ph}}^{32 \times 32}[\mathbf{I}_i] \quad T_{DCT_{ph}}^{16 \times 16}[\mathbf{I}_i]\}. \quad (13)$$

As shown in Figure 6, \mathbf{d}_i is a long ($4096 + 1024 + 256 = 5376$ dimensional) descriptor for each training image \mathbf{I}_i . Finally, we come up with a data matrix $\mathbf{D} = [\mathbf{d}_1 \quad \mathbf{d}_2 \quad \mathbf{d}_3 \quad \dots \quad \mathbf{d}_R]^T$ where each row contains a vector descriptor \mathbf{d}_i for exactly one training image. In order to extract only the relevant and least redundant features of this high dimensional feature matrix \mathbf{D} , a feature interaction-based dimensionality reduction strategy is used, as described in Subsection 3.5.

3.4.1. Obtaining invariance to occlusions

To show that the proposed multiscale feature extraction technique is invariant to occlusions, let’s pick a small subimage (called \mathbf{I}_{sub}^{new}) from the boosted image \mathbf{I} containing some occlusion. Irrespective of the noise source (but assuming linearity), \mathbf{I}_{sub}^{new} can mathematically be decomposed into the original visible subimage \mathbf{I}_{sub} plus a noisy subimage \mathbf{I}_{sub}^{noise} , i.e.:

$$\mathbf{I}_{sub}^{new} = \mathbf{I}_{sub} + \mathbf{I}_{sub}^{noise}. \quad (14)$$

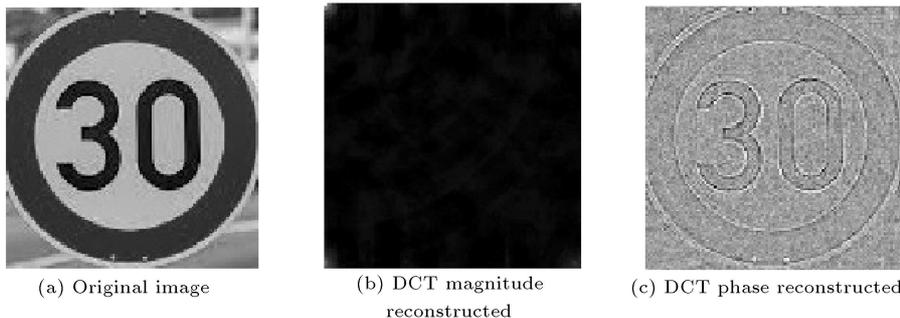


Figure 5. An illustration of discrete cosine transformed magnitude and phase reconstruction.

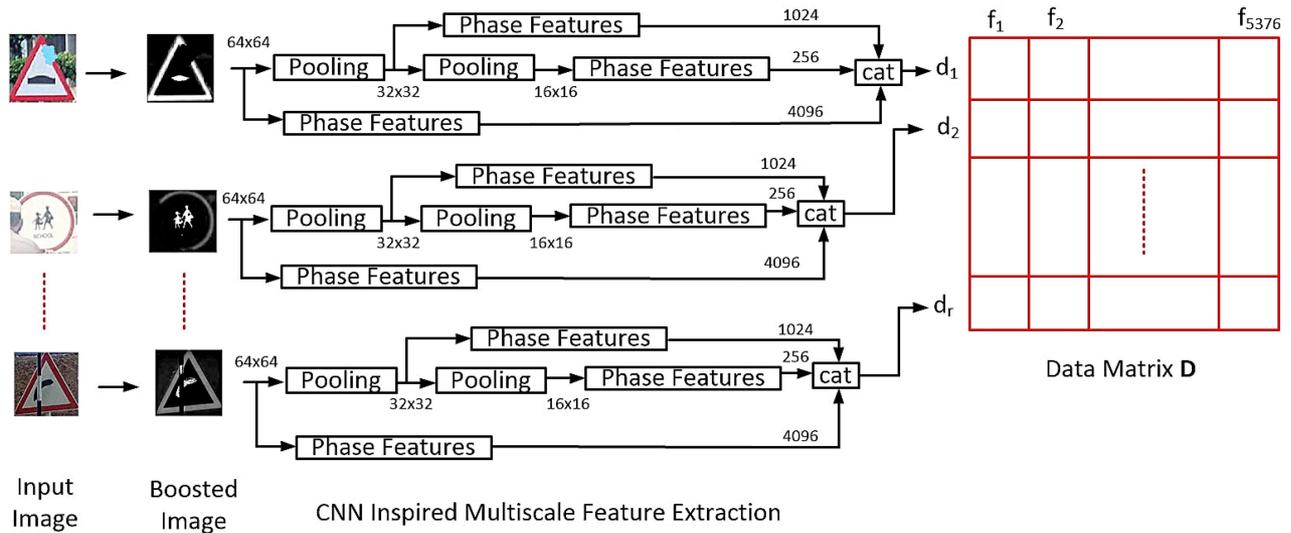


Figure 6. Description of feature extraction applied to training images and generation of data matrix D .

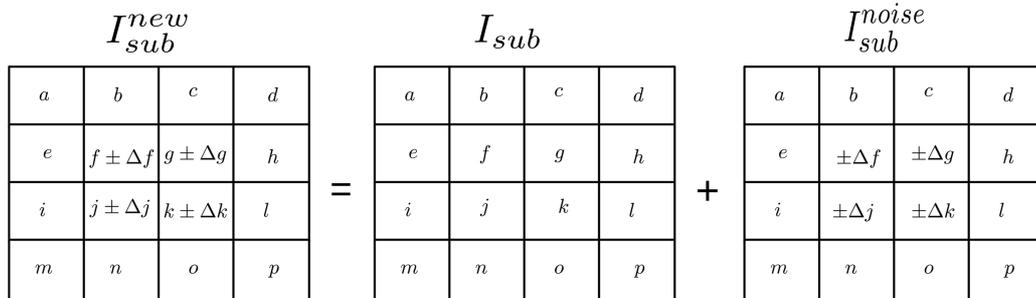


Figure 7. An occluded or degraded image represented as a sum of the original image and the noise.

A pictorial representation of the situation is given in Figure 7. At multiscale level $n \times n$, the descriptor computed using our proposed method for subimage \mathbf{I}_{sub}^{new} at position (x, y) can be extracted as [47]:

$$\begin{aligned} & \text{sign} \left(T_{DCT_{ph}}^{n \times n} [i_{sub}^{new}(x, y)] \right) \\ &= \text{sign} \left\{ T_{DCT_{ph}}^{n \times n} [i_{sub}(x, y)] + T_{DCT_{ph}}^{n \times n} [i_{sub}^{noise}(x, y)] \right\}. \end{aligned} \quad (15)$$

The operation is distributive because $T_{DCT_{ph}}^{n \times n}$ is a linear operator. For simplicity, $i_{sub}^{noise}(x, y)$, which is a pixel from noisy image (\mathbf{I}_{sub}^{noise}), is considered to be a random variable sampled from a one-dimensional Gaussian distribution with mean (μ_o) and variance (σ_0^2), i.e.:

$$i_{sub}^{noise}(x, y) \sim \mathcal{N}(\mu_o, \sigma_0^2). \quad (16)$$

The parameters are estimated as follows:

$$\mu_o = E [i_{sub}^{noise}(x, y)], \quad (17)$$

$$\sigma_0^2 = E [(i_{sub}^{noise}(x, y) - \mu_o)^2]. \quad (18)$$

As a result, an occluded or degraded pixel in image \mathbf{I}_{sub}^{new} can be written as:

$$i_{sub}^{new}(x, y) = i_{sub}(x, y) \pm K \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{[i_{sub}^{noise}(x, y) - \mu_o]^2}{2\sigma_0^2}}. \quad (19)$$

Here, K is an empirically chosen constant that determines the weight given to the random variable drawn from the Gaussian distribution $\mathcal{N}(\mu_o, \sigma_0^2)$. The scheme given in Eq. (19) can be used to simulate occlusions on a given sample traffic sign image. An example is given in Figure 8 where an occlusion is simulated on the top left corner of a SL30 traffic sign. Figure 8(a) gives a scenario where an improper value of K causes a greenish blur on the targeted area failing to introduce a significant effect of occlusion. In contrast, in Figure 8(b), a black patch over the desired area is visible which properly simulates the effect of occlusion.

Moreover, for any two scalars w and x , the sign of the sum is given by [48]:

$$\text{sign}(w + x) = \text{sign}(w) \quad \text{if } |w| > |x|. \quad (20)$$

Since $T_{DCT_{ph}}^{n \times n} [i_{sub}^{noise}(x, y)]$ is computed using a sparse

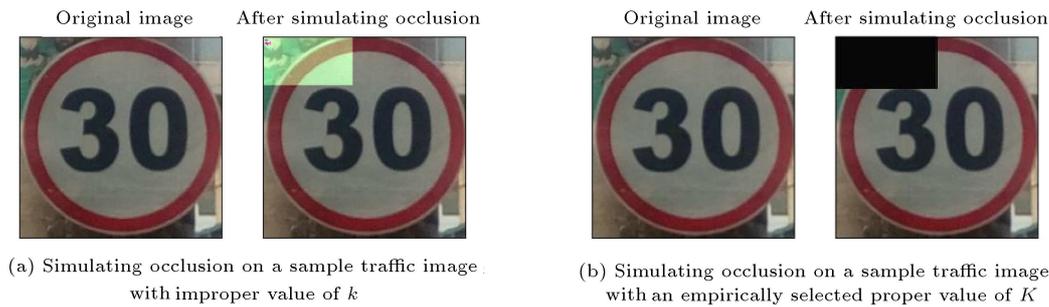


Figure 8. Simulating occlusion on a sample traffic sign image using improper and proper values of the constant K given in Eq. (19).

noisy image and obtained as a result of small positive or negative addition in the original subimage (refer to Figure 7), its magnitude is smaller than $T_{DCT_{ph}}^{n \times n}[i_{sub}(x, y)]$. Therefore, the right-hand side of Eq. (15) tends to retain the sign of $T_{DCT_{ph}}^{n \times n}[i_{sub}(x, y)]$ and hence, our proposed method remains stable in the presence of occlusions.

3.5. Feature interaction based dimensionality reduction

Since some information in the long feature vector matrix \mathbf{D} is more useful than the other, there is a need to select features that are strongly correlated with the target class (i.e., the contents of the sign, e.g., SL 100, Right Turn etc.) and are least redundant with respect to each other [49]. The operation is shown with the help of “Finding useful features” block in Figure 2. The high-dimensional feature matrix \mathbf{D} is fed to the dimensionality reduction operation and the output is a lower dimensional matrix \mathbf{S} of the most important features. As mentioned in Subsection 3.4, DCT phase information is the sign of the transform which is represented as a binary data, i.e., either +1 or -1. This facilitates the use of mutual information-based dimensionality reduction technique.

To obtain a subset of the most relevant and least redundant features, we propose the use of feature interaction-based dimensionality reduction [50,51]. Assuming columns of the data matrix \mathbf{D} are represented with \mathbf{f} , the main steps of the dimensionality reduction technique are given as follows:

1. $MI(\mathbf{f}; \mathbf{c})$ is mutual information based relevance between a class \mathbf{c} and a feature \mathbf{f} . For all features in the dataset \mathbf{D} , this procedure is repeated and finally the features are sorted in descending order with respect to their relevance to the target class;
2. The top ranked feature is the one that scores the highest class relevance. This feature is directly put into the subset of selected features \mathbf{S} without any comparison;
3. The second highest feature with respect to the class relevance (\mathbf{f}^+) is temporarily put in \mathbf{S} and mutual

information between this new subset $\{\mathbf{f}^+, \mathbf{S}\}$ and the class \mathbf{c} is computed, i.e., $MI(\mathbf{f}^+, \mathbf{S}; \mathbf{c})$;

4. Step 3 is repeated for all the remaining features in \mathbf{D} and the feature with the highest value of $MI(\mathbf{f}^+, \mathbf{S}; \mathbf{c})$ is added to the subset \mathbf{S} permanently;
5. In order to obtain required size of subset \mathbf{S} , the processes in Steps 3 and 4 are repeated equal to the number of required features.

An exact formula to compute joint mutual information (in Step 4) between candidate feature (\mathbf{f}^+) temporarily added to the already selected subset of features (\mathbf{S}) and the target class (\mathbf{c}) is given by Eq. (21). Right-hand side of the equation shows the same expression, but with the subset (\mathbf{S}) written in terms of its columns ($\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \dots, \mathbf{s}_s$). Each column of the subset \mathbf{S} expresses one feature:

$$MI(\mathbf{f}^+, \mathbf{S}; \mathbf{c}) = MI(\mathbf{f}^+, \mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \dots, \mathbf{s}_s; \mathbf{c}). \quad (21)$$

Since complex joint probability terms are involved in its direct computation which makes it computationally expensive, its approximation given in Eq. (22) is used, i.e., features are taken from \mathbf{S} one at a time; joint mutual information is computed; and the results are finally summed together [52].

$$MI(\mathbf{f}^+, \mathbf{S}; \mathbf{c}) \approx \sum_{\mathbf{s}_s \in \mathbf{S}} MI(\mathbf{f}^+, \mathbf{s}_s; \mathbf{c}). \quad (22)$$

Each term on the right-hand side of Eq. (22) can be expanded and written in the form of its joint and marginal entropy equivalents as given by Eq. (23). $H(\mathbf{f}^+, \mathbf{f}_s)$ represents the joint entropy term between the temporarily added feature and a sample feature picked from \mathbf{S} , $H(\mathbf{c})$ is the entropy of class label, and $H(\mathbf{f}^+, \mathbf{s}_s, \mathbf{c})$ is another joint entropy term.

$$MI(\mathbf{f}^+, \mathbf{s}_s; \mathbf{c}) = H(\mathbf{f}^+, \mathbf{s}_s) + H(\mathbf{c}) - H(\mathbf{f}^+, \mathbf{s}_s, \mathbf{c}). \quad (23)$$

It is noteworthy that since $H(\mathbf{c})$, the entropy of class label, remains the same throughout the feature selection procedure, it can be dropped; instead, a new computational formula for approximate joint mutual

information (MI') is given in Eq. (24). This approximation was used in our occlusion invariant multiscale feature data matrix \mathbf{D} to obtain a subset of the most important features \mathbf{S} .

$$MI'(\mathbf{f}^+, \mathbf{s}_s; \mathbf{c}) = H(\mathbf{f}^+, \mathbf{s}_s) - H(\mathbf{f}^+, \mathbf{s}_s, \mathbf{c}). \quad (24)$$

4. Experiments

We tested our proposed method for recognition of occluded traffic signs on two datasets namely Dataset 1 and Dataset 2. Detailed description of each dataset will be given in the next section. After feature extraction and dimensionality reduction steps, a classifier was trained separately for both datasets. To compare our proposed method with other state-of-the-art and recent methods, various evaluation measures like precision, recall, and recognition accuracy were used. Performance enhancement in terms of computational complexity and execution time was also monitored for the dimensionality reduction algorithm employed by our proposed technique.

4.1. Dataset

There is no publicly available dataset of occluded and degraded traffic signs right now. We collected around 1,000 real-world images containing partially occluded or degraded traffic signs by traveling on N5 and M2 highways in Pakistan [53]. To complete the dataset, a large number of images were synthetically produced using commonly found occlusions, i.e., tree leaves, vehicle, buildings, and other installations. Dataset1

contains around 1,903 images divided into 18 classes. A second dataset (Dataset 2) is composed of 1,321 images in 26 classes and was compiled using naturally occluded samples taken from GTSRB dataset [54] and synthetically occluded or degraded examples. Table 1 contains detailed information about the datasets used for experimentation. All images were rescaled to 64×64 to facilitate multiscale spectral feature extraction. Figure 9 shows some samples from the two datasets used for experiments. This is a maiden work to use real-world occlusions for experiments; all previous attempts for occluded and degraded traffic signs have used synthetic images only.

4.2. Classifiers

To evaluate our proposed method on two datasets (i.e., Dataset 1 and Dataset 2), a multiclass SVM classifier [19] was trained on various reduced subsets of relevant and unique features present in data matrix \mathbf{S} first on 18 classes for Dataset 1 and then for 26 classes on Dataset 2. Details of the datasets are given in Table 1. We used LibSVM [55] implementation of SVM classifier with linear kernel and other default parameters for experimentation. The classifier module is written in C language and supports classification among multiple classes. A core i7 computer with 8 GB RAM was used for experimentation.

4.3. Evaluation procedure

To evaluate the effectiveness of our proposed method, the available data was divided such that 60% data was used for training and the remaining 40% was kept aside

Table 1. Details of the two datasets used for experimentation (SL stands for Speed Limit).

Dataset	Instances	Classes	Class names	Composition	Source
Dataset 1	1,903	18	SL20, SL30, SL50, SL70,	Originally occluded	Field survey
			SL90, SL100, SL120, Slow,	Originally degraded	Field survey
			Warning right, Warning left,	Synthetically occluded	Occlusions from field data
			Round about, Falling rocks,	Standard	Field survey
			School, Traffic signals,	Negative	Field survey
			Road works, Road narrow, Roads meeting, Pedestrian		
Dataset 2	1,321	26	SL20, SL30, SL50, SL60, SL70,	Originally occluded	GTSRB dataset
			SL80, SL100, Proceed straight,	Synthetically occluded	Occlusions from field data
			Turn left, Turn right,	Standard	GTSRB
			Pass on left, Pass on right,	Negative	Field survey & GTSRB
			Round about, Warning right,		
			Warning left, Danger point, Traffic signals, Road works, Road narrow, Pedestrian, Bicycle, Kids crossing, Speed bump, Slippery road, Road bend, Roads meeting		

If the two sample means are statistically the same, the new sample of observations is augmented to the \mathbf{A} matrix shown in Figure 2. For this purpose, the following Hotelling’s t^2 statistic [57] is computed:

$$t^2 = \frac{n_{\mathbf{A}}n_{\mathbf{A}_{aug}}}{n_{\mathbf{A}} + n_{\mathbf{A}_{aug}}} \left(\boldsymbol{\mu}_{\mathbf{A}} - \boldsymbol{\mu}_{\mathbf{A}_{aug}} \right)^T \boldsymbol{\Sigma}_p^{-1} \left(\boldsymbol{\mu}_{\mathbf{A}} - \boldsymbol{\mu}_{\mathbf{A}_{aug}} \right). \quad (29)$$

Here, $n_{\mathbf{A}}$ and $n_{\mathbf{A}_{aug}}$ are the total number of samples in sets \mathbf{A} and \mathbf{A}_{aug} , respectively, and $\boldsymbol{\Sigma}_p$ is pooled covariance computed from the covariance matrices of sets \mathbf{A} and in \mathbf{A}_{aug} using the following formula:

$$\boldsymbol{\Sigma}_p = (n_{\mathbf{A}} - 1)\boldsymbol{\Sigma}_{\mathbf{A}} + (n_{\mathbf{A}_{aug}} - 1)\boldsymbol{\Sigma}_{\mathbf{A}_{aug}}. \quad (30)$$

The above hypothesis is tested at 95% confidence interval and if the two means $\boldsymbol{\mu}_{\mathbf{A}}$ and $\boldsymbol{\mu}_{\mathbf{A}_{aug}}$ are found to be statistically equal, the new \mathbf{A} matrix becomes:

$$\mathbf{A}_{t=t_1} = \left[\mathbf{A}_{t=t_0}^T \quad \mathbf{A}_{aug}^T \right]^T. \quad (31)$$

The system keeps on training itself dynamically and new estimates of mean and covariance are available after a number of test images are passed through the recognition step. This makes the system adaptive to new conditions and is able to continuously update itself in a newly added place anywhere in the world. Consequently, after introducing a number of samples with degraded images, the mean and covariance estimates are adjusted automatically to accommodate the new situation. A graphical illustration of the concept is shown in Figure 11; the system is trained on a set of standard

traffic signs at time $t = t_0$ initially; and then, the parameters keep on updating as it runs on test data. For the sake of understanding it better, a one-dimensional representation of the system is shown in Figure 12. The blue curve (drawn as a continuous line) shows parameter estimates corresponding to the RGB data (\mathbf{A}) available in the beginning, i.e., at $t = t_0$. The red dotted curve shows how it is updated at $t = t_1$ as new data (\mathbf{A}_{aug}) corresponding to the degraded signs is added at run time. This addition, when augmented to the original data, shifts both mean and variance estimates. Since the new data matrix $\mathbf{A}_{t=t_1}$ is of larger size than

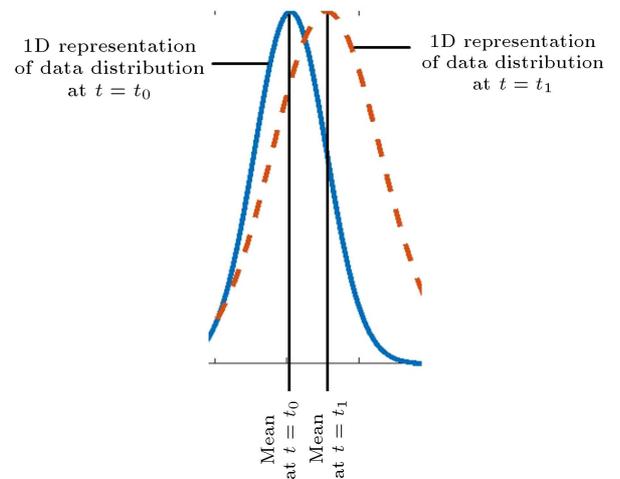


Figure 12. For the purpose of understanding, a one-dimensional visualization of the two-dimensional Gaussian distributions shown in Figure 11.

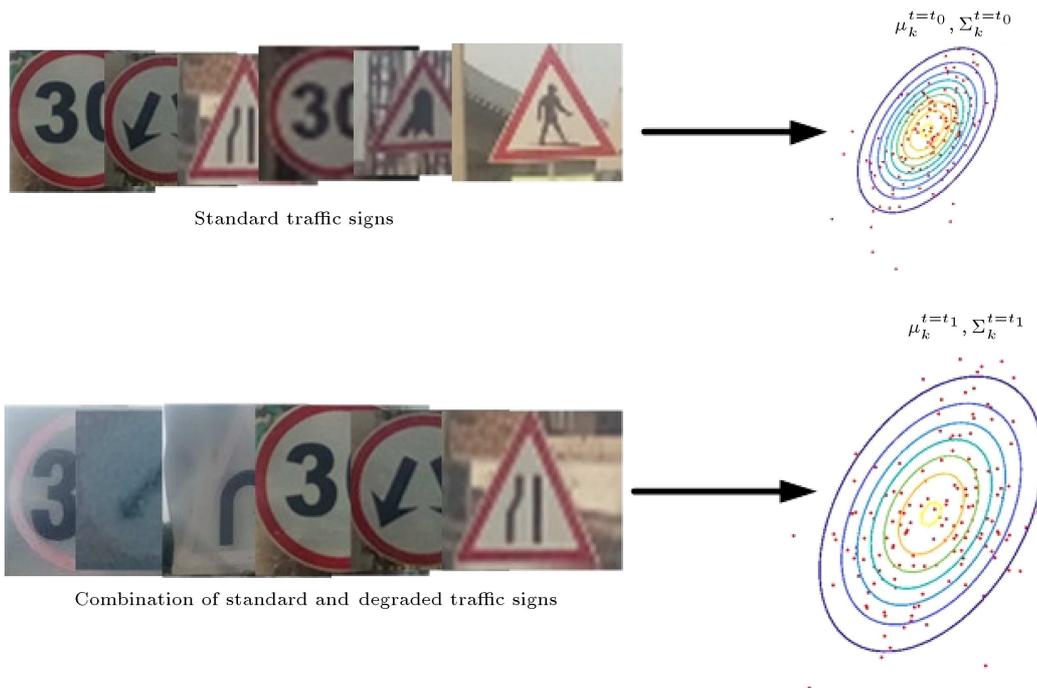


Figure 11. A visual representation of updates taking place after the system is run on a mixture of fine and degraded signs.

the matrix $A_{t=t_0}$, only the last z entries in $A_{t=t_1}$ matrix are used to update parameters where $z = size(A_{t=t_0})$.

5. Results

5.1. Comparing feature extraction techniques

In this section, we compare a number of feature extraction techniques with our proposed multiscale phase spectrum-based method. Considering precision and recall as measures, Table 2 shows a comparison of our proposed method with recent and state-of-the-art techniques previously used for recognition of visible and/or occluded traffic signs. A recent method based on occlusion maps presented by Hou et al. [27] and two other methods, i.e., Eigen [26] and energy [35] based methods, were found partially useful for the recognition of occluded and degraded traffic signs. However, novel feature extraction techniques such as HOG and LBP as well as their combination performed poorly for occluded traffic signs. However, our proposed technique obtains the highest precision and recall for both datasets.

As mentioned in Section 1, we divided the commonly found occlusions into four categories, namely Natural, Incidental, Deliberate, and Degradations. Table 3 shows recognition accuracy of these four categories of occlusions mentioned versus recent and novel methods for both datasets.

Our proposed method outperformed the other four methods by being the most accurate on both datasets. Moreover, the method was found to be the most effective for all categories of occlusions. Eigen space [26] and deformable models [35] were found better than occlusion maps based technique [27]. This is because the occlusion maps based method [27] was proposed mainly for detection of traffic signs but was found partially successful by the same authors for recognition too. An ant colony optimization-based method [49] performed well for natural category of occlusions for both datasets but produced poor results on other types of occlusions. Since this method employs an ant colony optimization-based feature selection technique, it can be thought of as the closest to our proposed techniques. However, this method was designed for standard traffic signs, whereas our proposed work targets traffic signs partially occluded by either other nearby objects or suffered with degradations.

5.2. Effect of dimensionality reduction

In order to know whether the feature interaction-based dimensionality reduction technique proposed in Subsection 3.5 is effective in enhancing recognition accuracy and reducing processing time, the whole process described in Section 3 was repeated with and without dimensionality reduction. The results given in Table 4 show that by employing feature interaction-

Table 2. Comparison of “Precision” and “Recall” among different methods. In literature, HOG and LBP were used for visible signs only, whereas occlusion maps, Eigen, and energy based methods were found partially useful for occluded traffic signs.

	Dataset	HOG [14]	LBP [58]	HOG+LBP	Eigen [26]	Energy [35]	Occlusion maps [27]	Proposed
Precision	Dataset 1	0.61	0.60	0.65	0.67	0.59	0.51	0.81
	Dataset 2	0.63	0.59	0.52	0.49	0.78	0.66	0.79
Recall	Dataset 1	0.59	0.48	0.65	0.61	0.66	0.65	0.79
	Dataset 2	0.43	0.49	0.62	0.59	0.68	0.71	0.76

Table 3. Comparing “Recognition Accuracy” of various methods on both datasets.

Dataset	Category	Eigen space [26]	Deformable models [35]	Occlusion maps [27]	ACO-RSDR [49]	Proposed
Dataset 1	Natural	0.71	0.68	0.60	0.71	0.80
	Incidental	0.64	0.58	0.63	0.65	0.76
	Deliberate	0.49	0.55	0.45	0.56	0.71
	Degraded	0.70	0.69	0.54	0.66	0.81
Dataset 2	Natural	0.75	0.77	0.65	0.75	0.83
	Incidental	0.65	0.45	0.55	0.71	0.69
	Deliberate	0.48	0.52	0.51	0.65	0.69
	Degraded	0.67	0.55	0.61	0.46	0.86

Table 4. Effect of using reduced feature set as a result of dimensionality reduction technique used in Subsection 3.5 on recognition accuracy and processing time.

Dataset	Recognition accuracy (No. of features)		Time on S (on D) (ms)
	Full feature set	Reduced feature set	
Dataset 1	0.79 (5,376)	0.81 (1,885)	30 (370)
Dataset 2	0.76 (5,376)	0.80 (1,915)	36 (258)

Table 5. Comparing time taken per image frame (seconds) by various methods on both datasets.

Dataset	Eigen space	Deformable models	Occlusion maps	ACO-RSDR	Proposed
	[26]	[35]	[27]	[49]	
Dataset 1	1.2	1.5	1.4	.99	0.99
Dataset 2	1.1	1.3	1.5	1.5	1.0

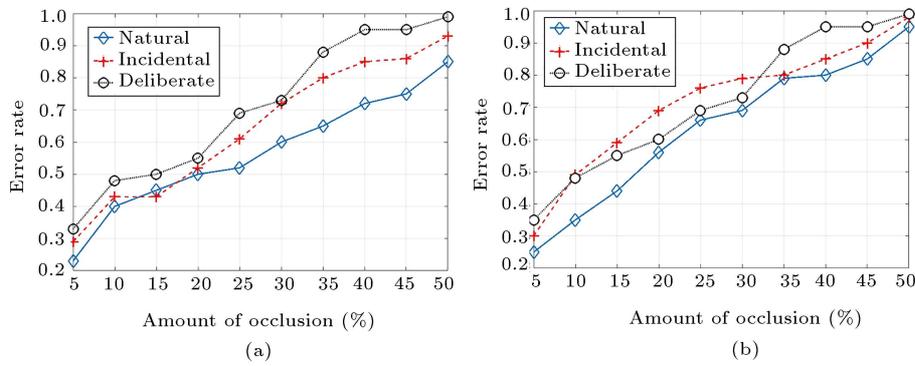


Figure 13. Average error rate of images from natural, incidental, and deliberate categories for (a) Dataset 1 and (b) dataset 2.

based dimensionality reduction, we achieved better average recognition accuracy on occluded and degraded traffic signs found in both datasets. Further, working on the data matrix of reduced features, i.e., S , saves computational time too. It is shown in Table 4 that with a reduced subset of only about 10% of features, we are able to obtain the accuracy achieved with 100% features. Moreover, more than 80% saving in execution time was also achieved for both datasets.

5.3. Amount of occlusion versus average error rate

It is obvious that increasing the number of occlusions causes error rate to increase. We performed experiments such that the degree of occlusion applied to traffic signs was increased from 5% to 50% (by synthetically putting real-world occlusions on standard traffic signs) and the error rate was computed for every sample. The results exhibit an increasing trend in error rate for individual signs belonging to each category, i.e., natural, incidental, and deliberate. Finally, an average error rate was computed for each category of signs found in both datasets and is shown in Figure 13. It is obvious that increasing the percentage of occlusion increases error rate for all four categories of occlusions and for both datasets.

5.4. Computational complexity

The proposed technique was also compared with other state-of-the-art and recent traffic sign recognition methods in terms of computational complexity. For this comparison, we have selected four methods, three of which partially address the problem of partial occlusion in traffic sign recognition [26,27,35] and the fourth incorporates an ant colony optimization-based feature selection technique [49]. The comparison given in Table 5 shows that all four methods have comparable computational complexity. However, the proposed method works just better than all other techniques in addition to being the most accurate, as demonstrated in the results given in Tables 2 and 3.

6. Discussion of results

It was demonstrated in Section 5 that our proposed method outperformed various other techniques on both datasets by achieving maximum precision, recall, and recognition accuracy values. Mainly, this advantage in performance is due to:

1. The computation of a boosted image in which the effect of occlusion is suppressed;
2. Our proposed recognition method takes maximum



Figure 14. A number of (a) correctly and (b) incorrectly recognized images by our proposed method.

benefit by continuously updating the underlying statistical model even during testing;

3. The use of dimensionality reduction increases both computational complexity and the overall accuracy.

An advantage achieved by our proposed method in precision and recall values (as shown in Table 2) in comparison to various other techniques mainly results from the computation of the boosted image and the use of our novel CNN inspired DCT phase-based multiscale feature extraction technique. On the other hand, the other techniques utilize either grayscale or selected channel of the RGB input image. The highest recognition accuracy for the first three types of occlusions mentioned in Table 3 is by virtue of our proposed feature extraction technique. However, for the fourth type, i.e., degraded, this is due to adaptive computation of the boosted image.

Further, the dimensionality reduction algorithm proposed in this work involves computing mutual information and feature interaction. Both of these operations involve joint probability terms and are, hence, computationally expensive. Since our descriptor contains only the binary values (either +1 or -1), computing joint probability terms becomes easier. Further, these computations are to be performed only during training and the resulting advantage is reduced execution time for test set.

It is shown in Figure 13 that the error rate for all categories of occlusions increases versus increasing

amount of occlusion, but the natural category is affected the least. This is because for this category, both color and texture of pixels belonging to occlusions differ from that of traffic signs, i.e., most of the plants are green and the repeating patterns (textures) on these occlusions have a non-uniform distribution of pixels. It is well known that for traffic signs, colors are bright and their distribution throughout the traffic sign is uniform. This characteristic of natural category of occlusion makes it the least prone to error. Occlusions of incidental and deliberate types are more likely to be affected by increasing the number of occlusions. Not surprisingly, this behavior remains the same for both datasets.

Finally, a number of correctly and incorrectly recognized signs by our proposed method are given in Figure 14. It can be seen that the proposed method fails as the occlusions cover more than 50% or the contents are completely suppressed.

7. Conclusion and future work

This study proposed a strategy to recognize occluded traffic signs. The task is challenging because a part of the traffic sign is suppressed by another object or the sign is weird or degraded. In this regard, different categories of occlusions found in real-world cases, namely natural, incidental, deliberate, and degradations, were identified. Given a detected partially occluded traffic sign image, we first measured a boosted image by taking inspiration from human cognition, suppressing

any occluded or degraded portions, and highlighting only the visible areas pertaining to the traffic sign. The statistical model used to compute boosted image is dynamic, i.e., the parameters are continuously updated as the system is tested in the field. To extract invariant features, a Convolutional Neural Network (CNN) inspired multiscale spectral method utilizing phase spectrum of discrete cosine transform was proposed. To remove irrelevant and redundant information, an interaction-based dimensionality reduction method was employed to improve recognition accuracy and save execution time. Experiments were conducted on two datasets namely Dataset 1 and Dataset 2. Dataset 1 was composed of naturally occluded images collected from authors' country of residence, in addition to synthetically occluded samples. Dataset 2 was derived from famous German Traffic Sign Recognition Benchmark (GTSRB) and it contained originally as well as synthetically generated images. Experimental results on both datasets illustrated that our proposed method outperformed state-of-the-art methods on both datasets. Moreover, the proposed dimensionality reduction technique showed outstanding savings in time and demonstrated a significant increase in accuracy.

As a future work, we are planning to investigate the performance of CNN on datasets composed of occluded and degraded traffic signs. Up till now, CNN-based methods reported in the literature aim at detection and recognition of standard traffic signs only. Since training a deep CNN is computationally expensive and is prone to overfitting for small datasets comprising only of a few thousand images, use of transfer learning to handle occlusions can be a possible future work. Another future effort is to test the adaptive parameter estimation in real-world scenarios for a longer duration of time and to analyze the behavior in case the amount of field/test data exceeds the initial training data.

References

- Gudigar, A., Chokkadi, S., and Raghavendra, U. "A review on automatic detection and recognition of traffic sign", *Multimedia Tools and Applications*, **75**(1), pp. 333–364 (2016).
- Noon, S., Javed, K., Mannan, A., et al. "Recognizing traffic signs using flexible discrete cosine transform (dct) grid", *Scientia Iranica, Transaction D, Computer Science & Electrical Engineering*, **24**(3), pp. 1384–1394 (2017).
- Bayouhd, K., Hamdaoui, F., and Mtibaa, A. "Transfer learning based hybrid 2d-3d cnn for traffic sign recognition and semantic road detection applied in advanced driver assistance systems", *Applied Intelligence*, **51**(1), pp. 124–142 (2021).
- Fan, B.B. and Yang, H. "Multi-scale traffic sign detection model with attention", *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, **235**(2–3), pp. 708–720 (2021).
- Jin, Y., Fu, Y., Wang, W., et al. "Multi-feature fusion and enhancement single shot detector for traffic sign recognition", *IEEE Access*, **8**, pp. 38931–38940 (2020).
- Cao, J., Zhang, J., and Huang, W. "Traffic sign detection and recognition using multi-scale fusion and prime sample attention", *IEEE Access*, **9**, pp. 3579–3591 (2020).
- Haque, W.A., Arefin, S., Shihavuddin, A., et al. "Deepthin: A novel lightweight cnn architecture for traffic sign recognition without gpu requirements", *Expert Systems with Applications*, **168**, p. 114481 (2021).
- Yazdan, R. and Varshosaz, M. "Improving traffic sign recognition results in urban areas by overcoming the impact of scale and rotation", *ISPRS Journal of Photogrammetry and Remote Sensing*, **171**, pp. 18–35 (2021).
- Cao, J., Zhang, J., and Jin, X. "A traffic-sign detection algorithm based on improved sparse r-cnn", *IEEE Access*, **9**, pp. 122774–122788 (2021).
- Liu, Z., Qi, M., Shen, C., et al. "Cascade saccade machine learning network with hierarchical classes for traffic sign detection", *Sustainable Cities and Society*, **67**, p. 102700 (2021).
- Liu, C., Chang, F., Chen, Z., et al. "Fast traffic sign recognition via high-contrast region extraction and extended sparse representation", *IEEE Transactions on Intelligent Transportation Systems*, **17**(1), pp. 79–92 (2015).
- Yang, Y., Luo, H., Xu, H., et al. "Towards real-time traffic sign detection and classification", *IEEE Transactions on Intelligent Transportation Systems*, **17**(7), pp. 2022–2031 (2016).
- Wang, D., Hou, X., Xu, J., et al. "Traffic sign detection using a cascade method with fast feature extraction and saliency test", *IEEE Transactions on Intelligent Transportation Systems*, **18**(12), pp. 3290–3302 (2017).
- Greenhalgh, J. and Mirmehdi, M. "Real-time detection and recognition of road traffic signs", *IEEE Transactions on Intelligent Transportation Systems*, **13**(4), pp. 1498–1506 (2012).
- Liu, C., Chang, F., and Liu, C. "Occlusion-robust traffic sign detection via cascaded colour cubic feature", *IET Intelligent Transport Systems*, **10**(5), pp. 354–360 (2016).
- Gudigar, A., Chokkadi, S., Raghavendra, U., et al. "Local texture patterns for traffic sign recognition using higher order spectra", *Pattern Recognition Letters*, **94**, pp. 202–210 (2017).
- Peng, H., Long, F., and Ding, C. "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**(8), pp. 1226–1238 (2005).

18. Guyon, I. and Elisseeff, A. “An introduction to variable and feature selection”, *Journal of Machine Learning Research*, **3**(Mar), pp. 1157–1182 (2003).
19. Schölkopf, B. and Smola, A.J., *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, 2nd Ed., pp. 230–259, MIT Press, USA (2002).
20. Liu, C., Li, S., Chang, F., et al. “Supplemental boosting and cascaded convnet based transfer learning structure for fast traffic sign detection in unknown application scenes”, *Sensors*, **18**(7), p. 2386 (2018).
21. Lai, Y., Wang, N., Yang, Y., et al. “Traffic signs recognition and classification based on deep feature learning”, *International Conference on Pattern Recognition Applications and Methods (ICPRAM)*, Funchal, Portugal, pp. 622–629 (2018).
22. Goodfellow, I., Bengio, Y., Courville, A., et al. *Deep Learning*, 1st Ed., pp. 91–150, MIT Press, USA (2016).
23. Lee, S.G., Sung, Y., Kim, Y.G., et al. “Variations of alexnet and googlenet to improve Korean character recognition performance”, *Journal of Information Processing Systems*, **14**(1), pp. 205–217 (2018).
24. Rosario, G., Sonderman, T., and Zhu, X. “Deep transfer learning for traffic sign recognition”, *IEEE International Conference on Information Reuse and Integration (IRI)*, Salt Lake City, Utah, pp. 178–185 (2018).
25. Tian, J. and Li, Y. “Convolutional neural networks for steganalysis via transfer learning”, *International Journal of Pattern Recognition and Artificial Intelligence*, **33**(2), p. 1959006 (2018).
26. Fleyeh, H. and Davami, E. “Eigen-based traffic sign recognition”, *IET Intelligent Transport Systems*, **5**(3), pp. 190–196 (2011).
27. Hou, Y.L., Hao, X., and Chen, H. “A cognitively motivated method for classification of occluded traffic signs”, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, **47**(2), pp. 255–262 (2017).
28. Mannan, A., Javed, K., and Noon, S.K. “Statistical boosting: A preprocessing technique to enhance performance of machine learning and deep learning models on partially occluded traffic signs”, *IEEE 23rd International Multitopic Conference (INMIC)*, Bahawalpur, Pakistan, pp. 1–6 (2020).
29. Mogelmoose, A., Trivedi, M.M., and Moeslund, T.B. “Vision based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey”, *IEEE Transactions on Intelligent Transportation Systems*, **13**(4), pp. 1484–1497 (2012).
30. Wang, X., Han, T.X., and Yan, S. “An hog-lbp human detector with partial occlusion handling”, *12th International Conference on Computer Vision*, Kyoto, Japan, pp. 32–39 (2009).
31. Rehman, Y., Riaz, I., Fan, X., et al. “D-patches: effective traffic sign detection with occlusion handling”, *IET Computer Vision*, **11**(5), pp. 368–377 (2017).
32. Floros, G., Kyritsis, K., and Potamianos, G. “Database and baseline system for detecting degraded traffic signs in urban environments”, *5th European Workshop on Visual Information Processing (EUVIP)*, Paris, France, pp. 1–5 (2014).
33. Li, L. and Ma, G. “Recognition of degraded traffic sign symbols using pnn and combined blur and affine invariants”, *Fourth International Conference on Natural Computation*, Jinan, China, pp. 515–520 (2008).
34. Ishida, H., Takahashi, T., Ide, I., et al. “Identification of degraded traffic sign symbols by a generative learning method”, *18th International Conference on Pattern Recognition (ICPR)*, Hong Kong, China, pp. 531–534 (2006).
35. De La Escalera, A., Armingol, J.M., Pastor, J.M., et al. “Visual sign information extraction and identification by deformable models for intelligent vehicles”, *IEEE Transactions on Intelligent Transportation Systems*, **5**(2), pp. 57–68 (2004).
36. Kim, J. “Detection of traffic signs based on eigen-color model and saliency model in driver assistance systems”, *International Journal of Automotive Technology*, **14**(3), pp. 429–439 (2013).
37. Tsai, L.W., Hsieh, J.W., Chuang, C.H., et al. “Road sign detection using eigen colour”, *IET Computer Vision*, **2**(3), pp. 164–177 (2008).
38. Ohta, Y.I., Kanade, T., and Sakai, T. “Color information for region segmentation”, *Computer Graphics and Image Processing*, **13**(3), pp. 222–241 (1980).
39. Alpaydin, E., *Introduction to Machine Learning*, 1st Ed., MIT Press, Istanbul, Turkey, pp. 57–103 (2004).
40. Maldonado-Bascón, S., Lafuente-Arroyo, S., Gil-Jimenez, P., et al. “Road-sign detection and recognition based on support vector machines”, *IEEE Transactions on Intelligent Transportation Systems*, **8**(2), pp. 264–278 (2007).
41. Dabbaghchian, S., Ghaemmaghami, M.P., and Aghagolzadeh, A. “Feature extraction using discrete cosine transform and discrimination power analysis with a face recognition technology”, *Pattern Recognition*, **43**(4), pp. 1431–1440 (2010).
42. Ayyalasmayajula, P., Grassi Pauletti, S., and Farine, P.A. “Retrieval of occluded images using dct phase and region merging”, *19th IEEE International Conference on Image Processing*, Orlando, FL, USA, pp. 2441–2444 (2012).
43. Chen, W., Er, M.J., and Wu, S. “Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain”, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, **36**(2), pp. 458–466 (2006).
44. Fracastoro, G., Fosson, S.M., and Magli, E. “Steerable discrete cosine transform”, *IEEE Transactions on Image Processing*, **26**(1), pp. 303–314 (2017).

45. Müller, J. “The hadamard multiplication theorem and applications in summability theory”, *Complex Variables and Elliptic Equations*, **18**(3–4), pp. 155–166 (1992).
46. Jia, H., Wang, W., Wang, D., et al. “Speech enhancement using modified mmse-lsa and phase reconstruction in voiced and unvoiced speech”, *International Journal of Pattern Recognition and Artificial Intelligence*, **33**(2), p. 1958002 (2018).
47. Gonzalez, R.C. and Woods, R.E., *Digital Image Processing*, 3rd Ed., Pearson Education, New Delhi, India, pp. 645–660 (2009).
48. Strang, G., Strang, G., Strang, G., et al., *Introduction to Linear Algebra*, 1st Ed., Cambridge Press, Wellesley, MA, pp. 60–160 (1993).
49. Jayaprakash, A. and KeziSelvaVijila, C. “Feature selection using ant colony optimization (aco) and road sign detection and recognition (rsdr) system”, *Cognitive Systems Research*, **58**, pp. 123–133 (2019).
50. Bannasar, M., Hicks, Y., and Setchi, R. “Feature selection using joint mutual information maximisation”, *Expert Systems with Applications*, **42**(22), pp. 8520–8532 (2015).
51. Brown, G., Pocock, A., Zhao, M.J., et al. “Conditional likelihood maximisation: a unifying framework for information theoretic feature selection”, *Journal of Machine Learning Research*, **13**(Jan), pp. 27–66 (2012).
52. Zeng, Z., Zhang, H., Zhang, R., et al. “A novel feature selection method considering feature interaction”, *Pattern Recognition*, **48**(8), pp. 2656–2666 (2015).
53. Shabbir, Y., Khokhar, M.F., Shaiganfar, R., et al. “Spatial variance and assessment of nitrogen dioxide pollution in major cities of Pakistan along n5-highway”, *Journal of Environmental Sciences*, **43**, pp. 4–14 (2016).
54. Wang, C.W. and You, W.H. “Boosting-svm: effective learning with reduced data dimension”, *Applied Intelligence*, **39**(3), pp. 465–474 (2013).
55. Chang, C.C. and Lin, C.J. “Libsvm: a library for support vector machines”, *ACM Transactions on Intelligent Systems and Technology*, **2**(3), p. 27 (2011).
56. Montgomery, D.C., Runger, G.C., and Hubele, N.F., *Engineering Statistics*, 5th Ed., John Wiley & Sons, New Delhi, India, pp. 33–89 (2009).
57. Mardia, K. “Assessment of multinormality and the robustness of hotelling’s t^2 test”, *Applied Statistics*, **24**(2), pp. 163–171 (1975).
58. Yuan, X., Hao, X., Chen, H., et al. “Robust traffic sign recognition based on color global and local oriented edge magnitude patterns”, *IEEE Transactions on Intelligent Transportation Systems*, **15**(4), pp. 1466–1477 (2014).

Biographies

Abdul Mannan received his BSc, MSc, and PhD degrees in Electrical Engineering from University of Engineering & Technology, Lahore, Pakistan in 2002, 2007, and 2021, respectively. Right now, he is working as an Assistant Professor at the Department of Electrical Engineering at NFC Institute of Engineering & Technology, Multan, Pakistan. His research interests include machine learning, pattern recognition, image processing, and computer vision.

Kashif Javed received the BSc, MSc, and PhD degrees in Electrical Engineering in 1999, 2004, and 2012, respectively, from the University of Engineering and Technology (UET), Lahore, Pakistan. He joined the Department of Electrical Engineering at UET in 1999 where he is currently a Professor. He has been a reviewer of IEEE Transactions on Knowledge and Data Engineering (TKDE) since 2011. His research interests include machine learning, pattern recognition, and natural language processing for Urdu.

Atta ur Rehman received BS degree in Electrical Engineering from Air university, Islamabad, Pakistan and PhD degree in the same field from Loughborough university, UK. Currently, he is with National University Science and Technology (NUST), Islamabad, Pakistan. His research interests are image processing, machine learning, and pattern recognition.

Haroon A. Babri received the BSc degree in Electrical Engineering from the University of Engineering and Technology (UET), Lahore, Pakistan in 1981 and the MSc and PhD degrees in Electrical Engineering from the University of Pennsylvania in 1991 and 1992, respectively. He was with the Nanyang Technological University, Singapore from 1992 to 1998, with the Kuwait University from 1998 to 2000, and with the Lahore University of Management Sciences (LUMS) from 2000 to 2004. He is currently a Professor of Electrical Engineering at UET. He has written two book chapters and has more than 60 publications in machine learning, pattern recognition, neural networks, and software reverse engineering.

Serosh K. Noon received her MSc degree in Electrical Engineering from University of Engineering and Technology, Lahore, Pakistan and is presently working towards her PhD degree in Electrical Engineering. She is also with the Department of Electrical Engineering at NFC Institute of Engineering and Technology, Multan, Pakistan. Her research interests include biomedical image and signal processing, artificial intelligence in agriculture, and computer vision.