

Gabor Filter and Gershgorin Disk-based Convolutional Filter Constraining for Image Classification

Vijay John, Ali Boyali, and Seiichi Mita

Abstract—In a standard convolutional neural network the different layers are trained using the back propagation algorithm. Following training, it can be observed that the first convolutional layer structurally resembles the Gabor filter in spite of being randomly initialized. Consequently, in this study, we propose to directly use the Gabor filters within the convolutional neural network's learning framework to enhance the classification accuracy. The Gabor filters are integrated into the learning framework in two ways. Firstly, they are used to initialize the filters in the first convolutional layer. Secondly, they are used in conjunction with the Gershgorin circle theorem to constrain the weights in the first convolutional layer during backpropagation. The Gabor filter-based training is investigated for its effectiveness in increasing the classification accuracy of the convolutional neural network. The proposed method is validated on the MNIST and USPS public dataset. The results obtained on these datasets show that integrating the Gabor filter within the neural network's learning framework increases the classification accuracy of the convolutional neural network.

Index Terms—Deep learning, gabor filters.

I. INTRODUCTION

Convolutional neural network (CNN) is a machine learning framework which has reported state-of-the-art results for various image classification [1] and nature language processing problems [2]. Unlike standard machine learning frameworks, the CNN framework, based on the multilayered perceptron, simultaneously performs feature extraction and classification. Problem specific features are extracted from the input, which are then classified. The extraction and classification are performed using multiple layers represented by weights and biases.

The multiple layers including the convolutional layer, the fully connected layer and the output layer are trained using the back-propagation-based gradient descent algorithm. In the standard CNN training framework, the CNN layer weights and biases are randomly initialized. Furthermore, no constraints are employed within the iterative back-propagation algorithm. At the end of the training in state-of-the-art networks such as the Alexnet [1], it can be observed that the structure of the weights in the first convolutional layer, typically, resembles the Gabor filter even with random initialization. Based on this observation, in this paper, we propose to directly utilize the Gabor filters directly within the CNN learning framework, as a means to enhance

the classification accuracy. The Gabor filters are used in two ways in the CNN framework. Firstly, they are used to initialize the weights in the first convolutional layer. Secondly, the Gabor filters are used to generate Gershgorin disks, which are then used as weight constraints within a modified backpropagation algorithm. The Gershgorin disks are defined by the Gershgorin circle theorem and are used to bound the spectrum of a square matrix. In this work, the Gershgorin disks derived from the square Gabor filters are used as weight constraints. More specifically, these constraints are used within a modified backpropagation algorithm to learn the weights of the first convolutional layer. We validate the proposed method on two public datasets, the MNIST [3] and USPS dataset. The results of our experimental validation shows that the proposed method improves the classification accuracy of the standard CNN-based classification framework. The main contribution of the proposed research is the integration of Gabor filter-based Gershgorin disks within the convolutional neural network's learning framework.

The remainder of the paper is structured as follows. Firstly, in Section II we review the literature. Next, we present the algorithm in Section III. The experimental results are presented in Section IV. Finally, we present our conclusions in Section V.

II. LITERATURE REVIEW

Convolutional neural network have been extensively in recent years for various classification and recognition applications [1], [3]. CNN achieves state-of-the-art results on various problem. But researchers strive to further improve the classification accuracy. One approach to increase the accuracy involves the used of different prior, such as the Gabor filter within the CNN framework [4], within the CNN framework. The Gabor filters are discriminative feature extractors which are based on the human vision. Apart from Gabor filters, researchers have also incorporated Log mel filter banks [5], [6], Fisher vectors [7] and sparse filter banks [8] within the CNN framework to improve the classification accuracy. The priors are either directly used as feature extraction [9] or used to initialize the feature extraction layers [10]. In the works by Kowlek *et al.* [9] and Mounika *et al.* [11], the Gabor filter are used to extract features from the input image in a pre-processing step. These features are then given as input to the neural network framework. Alternatively, in Calderon *et al.* [12], fixed Gabor filters are used in the first convolutional layer of Lenet-5 config, and the CNN training is restricted to the remaining layers. This approach is extended by Sarwar *et al.* [13], where the fixed Gabor filters

Manuscript received June 23, 2017; revised August 5, 2017.

V. John, A. Boyali, and S. Mita are with the Toyota Technological Institute, Nagoya, Japan (e-mail: {vijayjohn,boyali,smita}@toyota-ti.ac.jp).

are used in multiple layers. Both of these architectures demonstrate improved classification accuracy. Chang *et al.* [7] used the Gabor filters to initialize the first convolutional layer in a multi-layered CNN architecture. Subsequently, the complete network is fine-tuned to perform speech recognition, reporting improved classification accuracy. Additionally, the authors also show that initializing with the Gabor filter helps in faster convergence.

Compared to the literature, in our proposed work, we also propose to initialize the first convolutional layer using the Gabor filter. However, unlike the literature, we propose a Gabor filter and Gershgorin disk theorem-based algorithm to constrain the convolutional layer weights during the backpropagation fine-tuning.

III. ALGORITHM

Image classification is the research problem of estimating the class label l for a given input image I . In learning-based classification, like the CNN, models are trained with examples to perform the classification. The CNN is a neural network framework containing multiple layers including the convolutional and fully connected layers. These layers are represented using randomly initialized weights and biases, which are then trained using the backpropagation algorithm. Typically, the randomly initialized weights are trained without any constraints. In certain trained CNN networks [1], the first convolutional layer resembles the Gabor filter even with random initialization. Based on this observation, we propose to directly utilize the Gabor filters within the CNN. By directly incorporating the Gabor filter, we propose to enhance the classification accuracy of the CNN framework. The Gabor filters are used to initialize the weights in the first convolutional layer. Furthermore, the Gabor filters and Gershgorin disk-based constraints are used within the backpropagation algorithm to constrain the convolutional layer weights. Next, we review the algorithm components, before presenting the proposed algorithm.

A. Algorithm Components

The algorithm components include the convolutional neural network, the Gabor filters and the Gershgorin central disk theorem.

1) Convolutional neural network

CNN is an end-to-end learning framework used in various image classification and recognition applications [1, 3]. Unlike standard learning frameworks, the CNN simultaneously performs feature extraction and feature classification. CNN-based image classification is based on the multi-layered perceptron, where the initial layers perform feature extraction and the final layers perform the classification. The initial layers are represented by the convolution and pooling layers. These layers perform the feature extraction from the input image. The lower convolutional layers extract the low-level features, while the higher convolutional layers extract the high-level features. The convolutional layers are represented by learnable weights and biases, which are used to extract features from the preceding layers by convolution. A weight sharing strategy is

used in the convolutional layers, where the same learnable weights are convolved across the preceding layers. The extracted features are represented in the form of feature maps. The pooling layers function as sub-sampling layers. Unlike the initial layers, the final layers are represented by the fully connected layers, drop-out layers and the output layers. The learnable CNN weights and biases are learnt using the backpropagation algorithm with stochastic gradient.

In our work, since we want to evaluate the constrained learning of the convolutional filters, we choose a shallow CNN network. The three layered network contains CNN with one convolutional layer, one pooling layer, and one fully connected layer. The architecture for this canonical CNN network is as follows,

- Single channel input layer
- Convolutional layer with 12 filters of size 5×5 with hyperbolic tangent activation function.
- Pooling layer with average pooling with stride 2×2
- Fully connected output layer with 10 neurons and hyperbolic tangent activation function.

2) Gabor filtering

Gabor filters are linear filters used in image processing for feature extraction [14]. Gabor filters are bandpass filters which are convolved with the input image to extract the image features. The 2D Gabor filter are represented by a Gaussian kernel modulated with a sinusoidal plane wave. The real and imaginary 2D Gabor filters are given by,

$$G_c[i, j] = B e^{-\frac{(i^2+j^2)}{2\sigma^2}} \cos(2\pi f(i \cos \theta + j \sin \theta)) \quad (1)$$

$$G_s[i, j] = C e^{-\frac{(i^2+j^2)}{2\sigma^2}} \sin(2\pi f(i \cos \theta + j \sin \theta)) \quad (2)$$

where B and C are normalizing factors for the real and imaginary part, f is the frequency, θ is the orientation σ is the scale. By defining the parameters, the Gabor filter can extract the corresponding features from the input image.

3) Gershgorin central disk theorem

The Gershgorin circle theorem is used to generate the bound for the spectrum or the set of eigenvalues for a square matrix [15]. The spectrum bound is defined in the complex plane.

For a $K \times K$, matrix A , with elements a_{kj} , firstly, K closed Gershgorin disks $\mathbf{U}(\mathbf{a}, \mathbf{R}) = \{U(a_{kk}, R_k)\}_{k=1}^K$ are generated. Each k -th closed disk is centered on the the diagonal elements of the matrix A . The radius of the n -th disk is given by,

$$R_k = \sum_{j \neq k} |a_{kj}| \quad (3)$$

R_k is obtained by summing the absolute values of the non-diagonal entries of each row. Given the calculated Gerschgorin Disks, the circle theorem states that all the eigenvalues z of A lie within atleast one disk. The theorem is also described for disjoint union of disks, where an union of P disks could be disjoint from the union of the remaining $K-P$ disks.

In such a case, the theorem states that the union of P disks

contains P eigen values and the remaining eigenvalues are present in the union of $K-P$ disks. By creating a bound over the spectrum of A , their Eigenvalues can be calculated.

B. Proposed Algorithm

In the proposed algorithm, the Gabor filters are used to learn the first convolutional layer weights in the CNN. We next review the bootstrapping phase and modified backpropagation algorithm.

1) Weight initialization

In trained CNN-based image classifiers, like the Alexnet [1], the set of N randomly initialized filters of the first convolutional layer often resembles a Gabor filter. Based on this observation, we propose to directly initialize the first convolutional layer with the real part of N pre-defined filters $\mathbf{G} = \{G_c^n\}_{n=1}^N$. The Gabor filter-based initialization is performed to increase the rate of convergence and the classification accuracy. Note that we only select the real part of the Gabor filter to reduce the number of filters in the first convolutional layer.

2) Union-of-Gershgorin Disks

For the Gabor-based n -th convolutional weight, W_{c1}^n , with elements w_{kj} and size $K \times K$, the Gershgorin central disk theorem is used to generate the set of Gabor Gershgorin disks $\{U^n(w_{kk}, R_k)\}_{k=1}^K$. A set of Gabor Gershgorin disks is obtained as each k -th row in W_{c1}^n generates a corresponding Gabor Gershgorin disk $U^n(w_{kk}, R_k)$.

From the estimated set of K disks, all possible union-of-disks (UOD) is identified. The m -th union-of-disks corresponds to the union of p overlapping Gabor Gershgorin disks. The m different union-of-disks is represented as a set of M disjointed UOD, represented as $\Delta = \{\delta\}_{m=1}^M$. For example, if there are 2 disjointed UOD, with the first union containing P disks, and second union containing $K-P$, then δ_1 is represented by the union of P disks and δ_2 is represented by the union of the remaining $K-P$ disks.

Subsequently, we generate an UOD membership vector $\mathbf{d} \in \mathbb{R}^{1 \times K}$ for each filter. In the UOD membership vector, the k -th index corresponds to the k -th row in W_{c1}^n or equivalently the corresponding k Gabor Gershgorin disk $U^n(w_{kk}, R_k)$. The value for the k -th index $\mathbf{d}(k)$ directly corresponds to the m -th label for the k -th Gabor disk or row in Δ .

An illustrative example of Δ^1 and \mathbf{d}^1 , for a 5×5 matrix W_{c1}^n , is given in Fig. 1. The UOD generates 2 disjointed unions, where the first UOD ($m=1$) is represented by the union of overlapping Gabor disks generated by rows 1, 2, 3, while the second UOD ($m=2$) is represented by the union of overlapping Gabor disks generated by rows 4, 5. For this example, the set of disjointed UOD is $\Delta = [\delta^1, \delta^2]$ where,

$$\delta^1 = [U^1(w_{11}, R_1) \cup U^1(w_{22}, R_2) \cup U^1(w_{33}, R_3)] \quad (4)$$

$$\delta^2 = [U^1(w_{44}, R_4) \cup U^1(w_{55}, R_5)]. \quad (5)$$

The UOD membership vector estimated row-wise from the

Gabor filter is given as $\mathbf{d}^1 = [1, 1, 1, 2, 2]$ since the disks corresponding to rows 1, 2, 3 are present in the first UOD, while disks corresponding to the remaining rows are present in the second UOD. Note that Δ and \mathbf{d} are estimated for each n -th filter in the first convolutional layer. An illustration of the bootstrapping phase is shown in Fig. 2.

3) Constrained Backpropagation Algorithm

The standard CNN convolutional layer, fully connected layer and the output layer weights and biases are trained using the backpropagation algorithm. In this paper, the standard gradient descent-based backpropagation algorithm is used to train the fully connected and output layers. While, a modified gradient descent-based backpropagation algorithm is used to train the convolutional layer. Firstly, we present the standard weight update equation for the convolutional layer, which is given as,

$$W_{c1}(t+1) = W_{c1}(t) - \eta \nabla L(W_{c1}(t)) \quad (6)$$

where $W_{c1}(t)$ and $W_{c1}(t+1)$ are the convolutional filters at iteration t and $t+1$. $\nabla L(W_{c1}(t))$ is the negative gradient of the weight at iteration t . η is the learning rate for the negative gradient and is set to 0.01.

In the modified backpropagation algorithm, the weight update for the convolutional layers, in Eqn. 7, is constrained using the generated Gabor Gershgorin disks. More specifically, the convolutional weights are constrained to be similar to the initialized Gabor-based weights. To constrain each individual weight or filter at each $t+1$ iteration, we first calculate a corresponding buffer weight $B_{c1} \in \mathbb{R}^{(K \times K)}$ using Eqn 7. The buffer weight calculation for an individual weight is given as,

$$B_{c1} = W_{c1}(t) - \eta \nabla L(W_{c1}(t)) \quad (7)$$

In the next step, a "row-wise" weight update strategy is used to update the individual weight $W_{c1}(t+1)$ from the buffer weight. To perform the weight update, the disjointed UOD set Δ and UOD membership vector \mathbf{d} generated in bootstrapping phase are used. To perform the weight update, firstly, each k -th row in B_{c1} is considered and its Gershgorin disk \mathbf{U}^B is computed. Next, the UOD label (m) for this row is identified from the corresponding UOD membership vector ($\mathbf{d}(k)$). This identified label is used to retrieve the relevant UOD given as,

$$\delta_m = [U^1(w_{11}, R_1) \cup U^1(w_{22}, R_2) \cup U^1(w_{pp}, R_p)] \quad (8)$$

Finally, we check if there is any overlap between the buffer Gershgorin disk \mathbf{U}^B and the retrieved UOD. Only if there is an overlap between the buffer and Gabor Gershgorin disks, do we perform the weight transfer for the k -th row. This is used to constrain the weights close to the Gabor initialization. The weight update is given as,

$$W_{c1}(t+1)(k,:) = \begin{cases} B_{c1}(k,:), & \text{if } \mathbf{U}^B \text{ overlaps } \delta_m \\ W_{c1}(t)(k,:), & \text{otherwise no overlap} \end{cases} \quad (9)$$

An overview of the constrained backpropagation algorithm

is shown in Fig. 3. The pseudo code of the modified backpropagation algorithm is given in Algorithm 1.

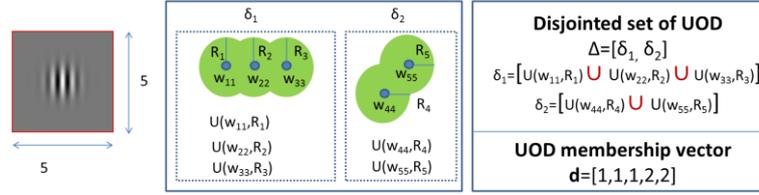


Fig. 1. An illustration of the disjointed set of UOD Δ and the UOD membership vector \mathbf{d} for a 5×5 Gabor filter.

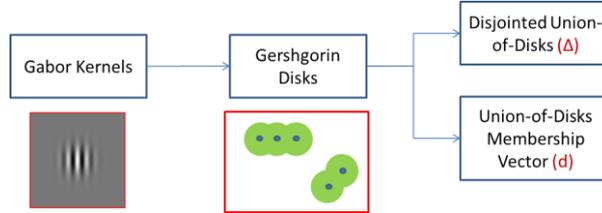


Fig. 2. An illustration of the Gabor initialization and the estimation of the Δ and \mathbf{d} in the bootstrapping phase.

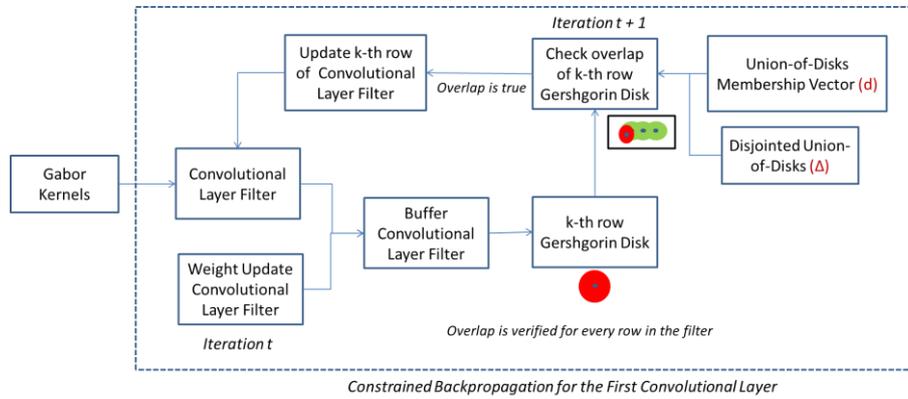


Fig. 3. An overview of the proposed constrained backpropagation algorithm. The iterative row-wise update of W_{c1} is shown.

IV. EXPERIMENTAL SECTION

We validate the proposed constrained backpropagation algorithm on the MNIST dataset and the USPS dataset. The MNIST and USPS dataset are image classification datasets used to classify digits. The MNIST dataset has 60000 training samples and 10000 testing samples, with each sample being a grey-scale image of size 28×28 . The USPS dataset contains 16×16 grayscale images with 7291 training observations and 2007 test observations.

TABLE I: CLASSIFICATION ACCURACY OF CNN-BASED DIGIT RECOGNITION ON THE MNIST – A DATASET WITH 10000 TEST SAMPLES

Algorithm	Wrong Det	Class. Acc.
CNN orig. backprop.	540	94.60
CNN constrained. backprop.	378	96.22

TABLE II: CLASSIFICATION ACCURACY OF CNN-BASED DIGIT RECOGNITION ON THE MNIST – B DATASET WITH 10000 TEST SAMPLES

Algorithm	Wrong Det	Class. accuracy
CNN orig. backprop.	222	97.78
CNN constrained. backprop.	211	97.89

TABLE III: CLASSIFICATION ACCURACY OF CNN-BASED DIGIT RECOGNITION ON THE USPS DATASET WITH 2007 TEST SAMPLES

Algorithm	Wrong Det	Class. accuracy
CNN orig. backprop.	140	93.02
CNN constrained. backprop.	134	93.32

Comparative analysis: We perform a comparative analysis

between the canonical CNN trained with gradient descent-based backpropagation algorithm and the proposed constrained backpropagation algorithm. We perform two sets of experiments with the MNIST dataset and one set of experiments with the USPS dataset. In the MNIST dataset, we perform a first experiment using 10000 training samples and 10000 testing samples. We refer to this experiment as *MNIST-a*, and perform the training over 10 epochs. The second MNIST experiment is performed using 60000 training samples and 10000 testing samples. We refer to this experiment as *MNIST-b*, and perform the training over 15 epochs. A single experiment is performed on the USPS dataset with 10 epochs.

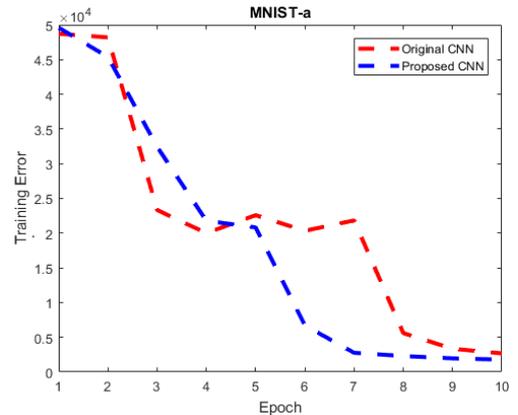


Fig. 4. Comparative analysis of the training errors for the *MNIST-a* dataset.

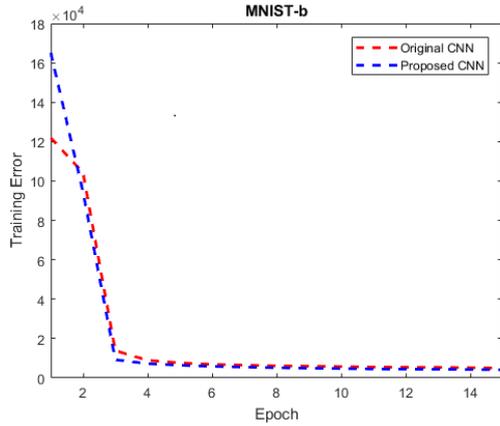


Fig. 5. Comparative analysis of the training errors for the *MNIST-b* dataset.

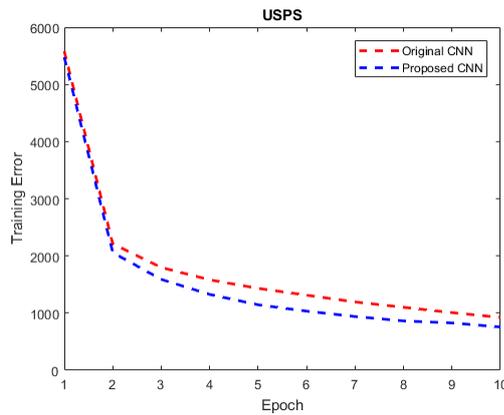


Fig. 6. Comparative analysis of the training errors for the *USPS* dataset.

Algorithm 1 Pseudo code of the Constrained backpropagation algorithm for each filter in the convolutional layer

$W_{c1}(t+1) = \text{Constrained Backpropagation}(B_{c1}, W_{c1}(t), \Delta, d)$

$W_{c1}(t+1) = W_{c1}(t)$

for each k -th row in B_{c1} do

 Generate Gershgorin disk U^B for the k -th row, $B_{c1}(k, 1:k)$

$\delta_m = \text{Retrieve UOD Index}(k, \Delta, d)$

$overlap = \text{Check Overlap}(U^B, \delta_m)$ if

$overlap$ then

$W_{c1}(t+1)(k,:) = B_{c1}(k,:)$.

 end if

end for

$\delta_m = \text{Retrieve UOD Index}(k, \Delta, d)$

Retrieve the UOD membership label, m , from the UOD membership vector $d(k)$

Retrieve the m -th UOD, δ_m , from the set of disjointed UOD Δ

Return δ_m

$overlap = \text{Check Overlap}(U^B, \delta_m)$

Retrieve the center-point, c_u , of the Gershgorin disk U^B

if c_u lies within the UOD δ_m then

$overlap = \text{true}$

 else

$overlap = \text{false}$

end if

Based on the experimental results as shown in Table I-Table III, the proposed algorithm improves the classification accuracy. Moreover, the training errors also improved, across the datasets (Fig. 4-Fig. 6). The results show us that by constraining the weights in the first convolutional layer using Gabor filter-based disks, the classification accuracy as well as the training efficiency is improved. On closer inspection of the results of the MNIST dataset, we see in case of larger training samples, *MNIST-b*, the results of the proposed

algorithm are marginally better than the original algorithm. On the other hand, with limited training samples, *MNIST-a*, the proposed algorithm is much better than the original algorithm.

V. CONCLUSION

In this work, the Gabor filters are directly incorporated within the convolutional neural network, as a means to enhance the classification accuracy. The Gabor filters are used to initialize the first convolutional layer. Additionally, Gershgorin disks derived from the Gabor filters constrain the weight update in the backpropagation algorithm. The advantages of constraining the weights are shown in the experiments. The algorithm is validated on the MNIST and USPS dataset, where we show that the proposed algorithm increases the classification accuracy constraining the backpropagation algorithm, the training error is reduced. In our future work, we will extend the Gershgorin disk-based constrain to colour image-based convolutional neural network. Additionally, we will also investigate the suitability of different filters for the initialization and constraints.

REFERENCES

- [1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, 2012.
- [2] B. Hu, Z. Lu, H. Li, and Q. Chen, "Convolutional neural network architectures for matching natural language sentences," *Advances in Neural Information Processing Systems*, vol. 27, 2014.
- [3] Y. LeCun and C. Cortes, "MNIST handwritten digit database," 2010.
- [4] J. G. Daugman, "Complete discrete 2D gabor transform by neural networks for image analysis and compression," *IEEE T. Acoust. Speech.*, vol. 36, no. 7, pp. 1169–1179, 1988.
- [5] S. Qu, J. Li, W. Dai, and S. Das, "Learning filter banks using deep learning for acoustic signals," *CoRR*, vol. abs/1611.09526, 2016.
- [6] T. N. Sainath, B. Kingsbury, A. R. Mohamed, and B. Ramabhadran, "Learning filter banks within a deep neural network framework," *ASRU*, 2013.
- [7] M. Cimpoi, S. Maji, and A. Vedaldi, "Deep convolutional filter banks for texture recognition and segmentation," *CoRR*, 2014.
- [8] L. Pfister and Y. Bresler, "Learning sparsifying filter banks," *SPIE*, 2015.
- [9] Kwolek, "Face detection using convolutional neural networks and gabor filters," *ICANN*, vol. 1, 2005.
- [10] S. Chang and N. Morgan, "Robust cnn-based speech recognition with gabor filter kernels," *International Speech Communication Association*, 2014.
- [11] B. Mounika, N. Reddy, and V. Reddy, "Neural network based face detection using gabor filter response," *International Journal of Neural Networks*, vol. 2, pp. 2249–2771, 2012.
- [12] A. Calderon, S. Roa, and J. Victorino, "Handwritten digit recognition using convolutional neural networks and gabor filters," in *Proc. International Congress on Computational Intelligence*, 2003.
- [13] S. S. Sarwar, P. Panda, and K. Roy, "Gabor filter assisted energy efficient fast learning convolutional neural networks," *ArXiv e-prints*, May 2017.
- [14] Fogel and D. Sagi, "Gabor filters as texture discriminator," *Biological Cybernetics*, vol. 61, no. 2, pp. 103–113, 1989.
- [15] R. S. Varga, *Gershgorin and His Circles*. Springer Series in Computational Mathematics, 2004.



Vijay John received the M. S degree in information technology and robotics from Carnegie Mellon University, USA in 2005. He received the Ph.D degree in computer vision from the University of Dundee, Dundee, UK in 2007. He then joined the University of Amsterdam in 2011 as a postdoctoral researcher, and was a visiting researcher at Philips Research, Netherlands during the summer of 2012. In 2014, he

joined the Toyota Technological Institute, Nagoya, Japan, as a postdoctoral research fellow. Since 2017, he is an assistant professor at the Toyota Technological Institute. His current research interests include computer vision, machine learning with applications in autonomous vehicles and human motion analysis.



Ali Boyali received his bachelor degree from the Mechanical Engineering Faculty of the Istanbul Technical University in 1998, M.Sc. degree in automotive engineering in 2002, and his Ph.D. in automotive mechatronics in 2008. He is now with the National Institute of Advanced Science and Technology in Japan as a post-doctoral research fellow. His recent researches focus on the development robotic wheelchairs that can be commanded by the recognized gesture and postures of the people with mobility impairment.



Seiichi Mita received the B.S. degree, the M.S. degree and the Ph.D. degree in electrical engineering from Kyoto University in 1969, 1971, 1989 respectively. He studied at Hitachi Central Research Laboratory, Kokubunji, Japan from 1971 to 1991 on signal processing and coding methods. He moved to Data Storage & Retrieval Systems Division, Hitachi, Ltd. in 1991. He developed channel coding methods and these LSI chips for magnetic disk drives. Now, he is a professor of Toyota Technological Institute (TTI) in Nagoya since 1999 and a director of Research Center for Smart Vehicles at TTI. Currently, he is greatly interested in the research area of autonomous vehicles and sensing systems. He is a member of the Institute of Electronics, Information and Communication Engineers and the Institute of Image Information and Television Engineers in Japan. He is also a member of IEEE. He received the best paper award of IEEE Consumer Electronics Society in 1986 and the best paper and author awards of the Institute of Television Engineers in Japan in 1987 and 1992 respectively.