Journal of Electronic Imaging

JElectronicImaging.org

Vehicle license plate recognition using visual attention model and deep learning

Di Zang
Zhenliang Chai
Junqi Zhang
Dongdong Zhang
Jiujun Cheng

SPIE. IS&T imaging.org

Vehicle license plate recognition using visual attention model and deep learning

Di Zang,[a,b,*] Zhenliang Chai,[a,b] Junqi Zhang,[a,b] Dongdong Zhang,[a,b] and Jiujun Cheng[a,b]
[a]Tongji University, Department of Computer Science, No. 4800, Cao'an Highway, Shanghai 201804, China
[b]Ministry of Education, The Key Laboratory of Embedded System and Service Computing, No. 4800, Cao'an Highway, Shanghai 201804, China

Abstract. A vehicle's license plate is the unique feature by which to identify each individual vehicle. As an important research area of an intelligent transportation system, the recognition of vehicle license plates has been investigated for some decades. An approach based on a visual attention model and deep learning is proposed to handle the problem of Chinese car license plate recognition for traffic videos. We first use a modified visual attention model to locate the license plate, and then the license plate is segmented into seven blocks using a projection method. Two classifiers, which combine the advantages of convolutional neural network-based feature learning and support vector machine for multichannel processing, are designed to recognize Chinese characters, numbers, and alphabet letters, respectively. Experimental results demonstrate that the presented method can achieve high recognition accuracy and works robustly even under the conditions of illumination change and noise contamination. [DOI: 10.1117/1.JEI.24.3.033001]

Keywords: license plate recognition; visual attention model; convolutional neural network; intelligent transportation system; deep learning.

Paper 15006 received Jan. 5, 2015; accepted for publication Apr. 2, 2015; published online May 4, 2015.

1 Introduction

In the past two decades, intelligent transportation systems have been developed to improve public transportation safety and mobility by integrating multiple advanced technologies. Automatic identification of vehicles has become more and more important in many applications; for example, parking fees and toll payments, traffic surveillance, ticket issuing, access control, and so on. A license plate is a unique feature by which to identify each individual vehicle. Automatic recognition of vehicle license plates, as an important research area of an intelligent transportation system, has already been widely studied for some decades. License plates may have different formats for different countries; however, the basic techniques to recognize them are the same, i.e., license plate detection, segmentation, and character recognition. In this paper, we aim to address the problem of Chinese car license plate recognition in traffic videos for civil use. As shown in Fig. 1, the Chinese car license plate consists of seven segments, where the left most segment is a Chinese character with 31 possible values indicating the region to which the car belongs. The remaining six segments of the license plate are either numbers or alphabet letters with a total of 34 possible values; the Chinese license plate excludes letters "O" and "I" because they look like numbers 0 and 1. There exist many research articles regarding car license plate recognition. The first crucial step is to detect the license plate. The localization accuracy can greatly affect the recognition rate. Due to the presence of dense edge sets, edge-based methods[1–7] are the most popular ways to localize the license plates. Texture[8,9] or the combinations of colors[10–12] are also considered as key features for license plate detection. Gendy et al.[13] and Yanamura et al.[14] have applied Hough transforms to detect the frames containing borders of license plates. In Ref. 15, a principal visual word is used to automatically locate license plates. As alternative ways, morphological methods[2,4] were proposed to segment license plates from original images. Kim et al.[16] and Zhang et al.[17] employed the genetic algorithm and AdaBoost learning algorithm, respectively, to recognize license plates. Neural network-based approaches[18–21] are also frequently used.

By selectively increasing the activity of sensory neurons that represent the relevant locations and features of the environment, visual attention[22] enables the visual system to process potentially important objects. Based on the investigation of visual characteristics, some researchers presented visual attention models[23–26] to detect objects of interest. The most popular visual attention mode is the one[25] proposed by Itti et al., which integrates multiple low-level features to generate a saliency map for object detection. It has been widely used in many applications[27–29] due to the competitive detection performance. In this paper, we propose a modified visual attention model to localize the position of Chinese license plates.

The second key step is character segmentation. The most popular ones are the projection method,[30–32] combined features,[33,34] and connection components.[3] The projection method is simple and fast, and allows characters to be segmented according to their height and width values once the frame containing the license plate boundaries is determined. In this paper, we employ the projection method to segment Chinese license plate characters.

Once the license plate is segmented into a couple of blocks, the last stage is to recognize the characters. Template matching methods[35,36] are simple and straightforward; however, they are vulnerable to any font, rotation, noise, and thickness

*Address all correspondence to: Di Zang, E-mail: zangdi@tongji.edu.cn

Fig. 1 Examples of Chinese car license plates and 31 Chinese characters.

change. Other popular approaches are artificial neural networks[18,37–39] and classifiers.[6,40,41]

Convolutional neural network (CNN) is one of the ways to perform deep learning, where raw images can be directly used as inputs. Because of its local receptive fields, shared weights and the spatial subsampling, CNN has the advantage of shift, scale, and noise distortion invariance. Due to the deep network structure, CNN is able to learn a hierarchy of features by building high-level features from low-level ones to describe objects. CNN was first employed for handwriting recognition[42] and it achieved a very high recognition rate. In Refs. 19 and 20, convolutional neural networks were used for license plate detection. Until now, CNN has been widely used for visual object recognition,[43] human action recognition,[44] brain-computer interaction,[45] and audio classification,[46] and the corresponding systems yield very competitive performances.

In order to obtain high recognition rates, in this paper, we propose two classifiers to recognize Chinese license plate segments by coupling CNN-based feature learning and support vector machine (SVM) into a single framework for multichannel processing.

2 Proposed Method

2.1 *System Architecture*

The system architecture of the proposed Chinese car license plate recognition is shown in Fig. 2. Given multiple frames of a traffic video, we first segment cars from original frames by computing the motion information in a region of interest. Then car images are preprocessed to remove noise and enhance image contrast. By fusing multiple features to generate a saliency map, a modified visual attention model is able to detect the position of the license plate.

Once the car license plate region is divided from the original car image, it can be further segmented into seven blocks. By integrating CNN-based feature extraction and SVM into a single framework for multichannel processing, two new classifiers are designed for recognizing Chinese characters, numbers, and alphabet letters, respectively. At this stage, CNNs and SVMs are previously trained with a training sample database.

2.2 *Chinese Car License Plate Detection Using Visual Attention Model*

Detecting license plates is the first key step for car license plate recognition. The detection accuracy significantly affects the performance of the whole system. Inspired by the traditional visual attention model,[25] we propose a modified visual attention model to detect the position of license plates as shown in Fig. 3.

For Chinese car license plates, a great many of them have a blue or yellow background, we thus use a color BY as the combination of blue and yellow for building the color feature map. Given a color image with red, green, and blue channels (r,g, and b), the color BY can be computed as $B-Y$, where B and Y are represented as the following:

$$B = b - \frac{r+g}{2} \quad Y = \frac{r+g}{2} - \frac{|r-g|}{2} - b. \tag{1}$$

The intensity of the car image is obtained as $I = r + g + b/3$, which is the average of three color channels. To extract the orientation features along lines of 0 deg, 30 deg, 60 deg, 90 deg, 120 deg, 150 deg, and 180 deg, the Gabor filter is employed.

Similar to the traditional visual attention model, the color, intensity, and orientation features are used to build Gaussian pyramids, and center-surround operations[25] are applied to generate corresponding color, brightness, and orientation feature maps.

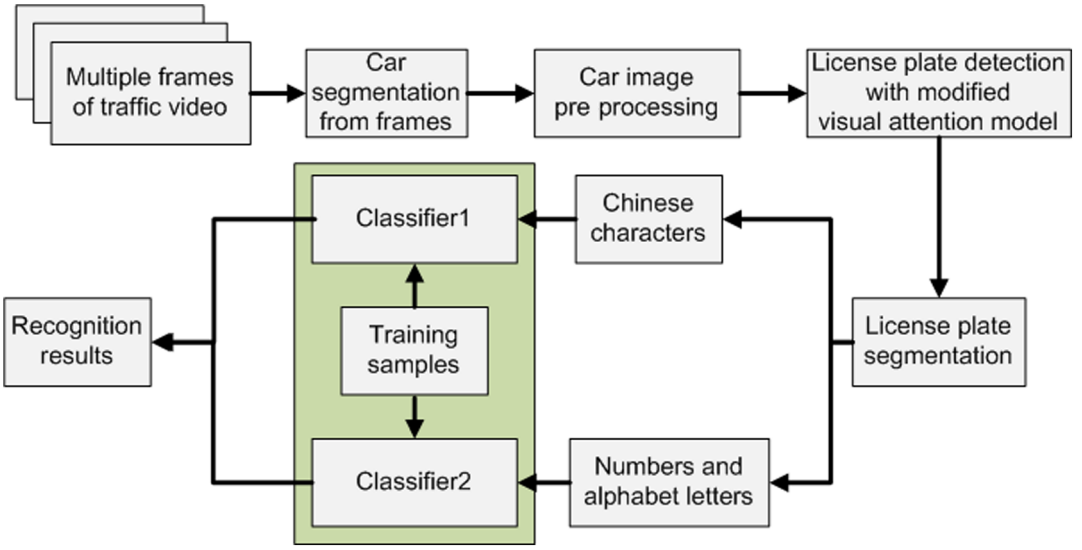

Fig. 2 System architecture of Chinese car license plate recognition in traffic videos.

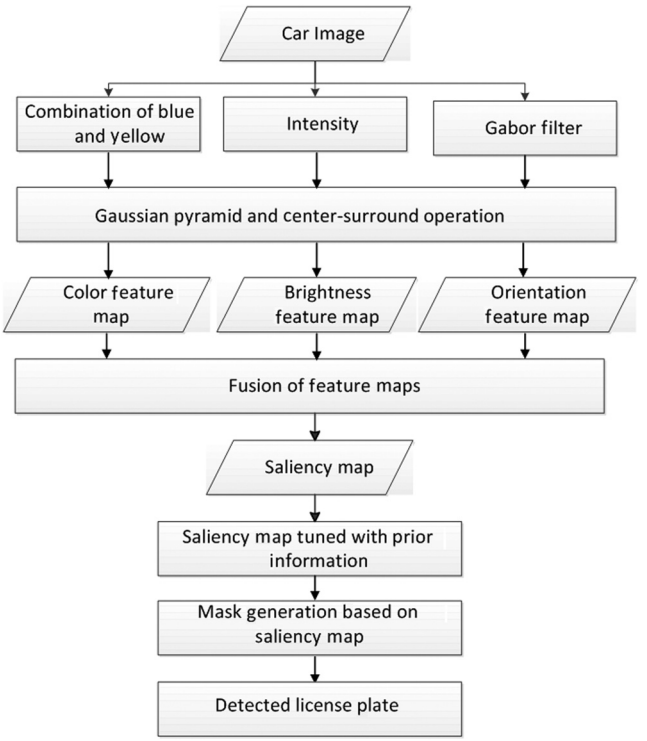

Fig. 3 Modified visual attention model for license plate detection.

Given an image $f(x,y)$, the Gaussian filtered result $G_\sigma(x,y)$ can be obtained as

$$G_\sigma(x,y) = f(x,y) * h_\sigma(x,y), \tag{2}$$

where $*$ means the convolution operator, and the Gaussian kernel $h_\sigma(x,y)$ with scale parameter σ reads

$$h_g(x,y) = \frac{1}{2\pi\sigma^2} e^{\left(-\frac{x^2+y^2}{2\sigma^2}\right)}. \tag{3}$$

The Gaussian pyramid is then constructed by progressively filtering the input image with Gaussian kernels and continuously downsampling the outputs by a factor of 2. The center-surround operation can be represented as

$$G_c \Theta G_s = G_c - \mathrm{Inter}_{s\to c}(G_s), \tag{4}$$

where Θ means the center-surround operator, c and s refer to low and high scale parameters, G_c and G_s are the images with lower and higher scales inside the Gaussian pyramid, and $\mathrm{Inter}_{s\to c}$ indicates that G_s is interpolated as the same size of G_c. This operation is used to compute the difference between fine and coarse scales. The center indicates a pixel at scale $c \in \{2,3,4\}$, and the surround is the corresponding pixel at scale $s = c + d$, with $d \in \{3,4\}$. In this way, 6 color maps, 6 intensity maps, and 42 orientation maps can be computed. Merging multiple images generated by the center-surround operators can yield a two-dimensional feature map S

$$S = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} N[G(c,s)], \tag{5}$$

where $G(c,s) = G_c \Theta G_s$ is the output from the center-surround operation, N refers to the normalization that is similar to Ref. 25, and $\bigoplus$ means adding images pixel by pixel. The feature map can thus be considered as the combination of images $G(2,5)$, $G(2,6)$, $G(3,5)$, $G(3,6)$, $G(4,5)$ and $G(4,6)$. Given a car image, we are able to build the corresponding color, intensity, and orientation feature maps as shown below:

$$\begin{aligned} S_C &= \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} N[BY(c,s)] \\ S_I &= \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} N[I(c,s)] \\ S_O &= \sum_{\theta} N\{\bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} N[O(c,s,\theta)]\}, \end{aligned} \tag{6}$$

where $\theta \in \{0\ \mathrm{deg}, 30\ \mathrm{deg}, 60\ \mathrm{deg}, 90\ \mathrm{deg}, 120\ \mathrm{deg}, 150\ \mathrm{deg}, 180\ \mathrm{deg}\}$. To localize the car license plate, color, intensity, and orientation feature maps are fused to generate a saliency map S_{map}

$$S_{\mathrm{map}} = w_1 S_C + w_2 S_I + w_3 S_O, \tag{7}$$

where $w_1 + w_2 + w_3 = 1$ and w_1, w_2, and w_3 indicate different weights associated with color, intensity, and orientation features, respectively.

Fusing color, intensity, and orientation feature maps can yield a saliency map, where regions with values above the threshold indicate high visual attention and are candidate positions for the license plate. In this paper, Ostu's automatic thresholding method[47] is applied to adaptively decide the thresholds. Even though the initial map may contain more than one such region, the saliency map for license plates can be finely tuned according to the prior information. Generally, the license plate in the car image is located in the bottom and middle areas. We divide the car image into 12 blocks with 4 rows and 3 columns, so the middle block in the bottom row is considered as the region of interest to remove unrelated areas in the saliency map. Furthermore, the ratio between the length and width of the license plate, and the rectangle shape are also used as constraints to decide the plate region in the saliency map.

Based on the tuned saliency map, a binary mask image is generated by assigning ones to the corresponding pixels whose values are greater than the threshold, where thresholding values are also determined using Ostu's method.[47] The license plate can thus be separated by combining the original car image and the mask image.

2.3 *License Plate Segmentation Using Projection Method*

When the license plate is detected and separated from the car image, the projection method[30–32] can be applied to segment the license plate into seven blocks. As shown in Fig. 4, the license plate image should first be transformed into a binary image, where characters are white and the background is black. Then, white pixels are projected both horizontally and vertically. According to the horizontal projection, the

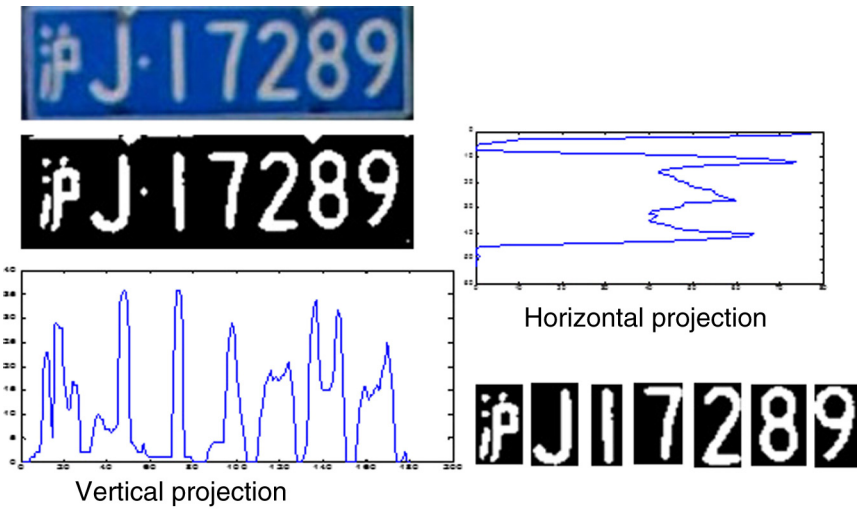

Fig. 4 License plate segmentation using the projection method.

top and bottom boundaries of characters can be found to compute the approximate height.

By checking zero points of the vertical projection curve, some regions can be segmented. According to the camera setup configuration, every segment usually has a pixel width between 15 and 25, thus regions with a pixel width in this range are chosen to compute an average width for cutting the license plate into seven segments, and the most left segment contains the Chinese character.

2.4 *Chinese Car License Plate Recognition Based on Deep Learning*

Deep learning is a new technique in the area of machine learning, which attempts to model high-level abstractions in data. There are various deep learning architectures such as deep convolutional neural network, deep belief network, and so on. Due to the deep network structures, they have been widely used in many applications with great success. Compared with other architectures, a deep convolutional neural network has fewer weights and less complexity. In addition, it allows one to directly use original images as inputs which enables a hierarchical learning of features. A classical convolutional neural network is robust against image distortion and affine transformation, but it cannot always produce optimal classification results. By finding a hyperplane in the feature space, SVMs are able to yield good classification while maximizing the margin in such a space. Therefore, we integrate CNN-based feature extraction and SVM into a single framework to design classifiers for license segments recognition.

Figure 5 shows the structure of classifier 1 for identifying the Chinese character. Given a segmented license block image, it is first decomposed into red, green, and blue channels. Images in each channel are then used as inputs for CNNs to hierarchically learn features and the results are delivered to SVMs for generating target label probability values.

Since there are a total of 31 different Chinese characters, the output layer of the SVM is set to have 31 nodes. Outputs from three SVMs are then fed to a majority voting process to make the final decision. The SVM can output probability values for each class. If the highest probability values in two or three channels indicate the same class, this class will be selected as the final recognition result. However, when the highest probability values correspond to three different classes, we compute the average of the SVM probability vectors for three channels and choose the class associated with the highest value as the final result.

The second classifier to recognize numbers and alphabet letters has a very similar structure. However, the inputs are either images of numbers or alphabet letters and the SVM in this classifier contains 34 nodes at the output layer.

In this paper, CNNs have the same architecture. In contrast to the traditional one in Ref. 42, the used CNN has a network structure including one input layer, three convolutional layers with different kernel sizes and three subsampling layers, as shown in Fig. 6.

The input of the used CNN is a normalized image with a size of 38×38 pixels. The first convolutional layer $C1$ consists of eight feature maps, and every feature map corresponds to 32×32 neurons. Each neuron in the feature map is connected to its corresponding 7×7 receptive field of the input image as indicated by the red square. Weights associated with the 7×7 connections are pretrained and shared inside every feature map, thus they are considered as the coefficients of the convolution kernel. In such case, each feature map can be regarded as the convolution result. Sharing the same 7×7 connection weights in each feature map enables CNN to be invariant to shift and rotation

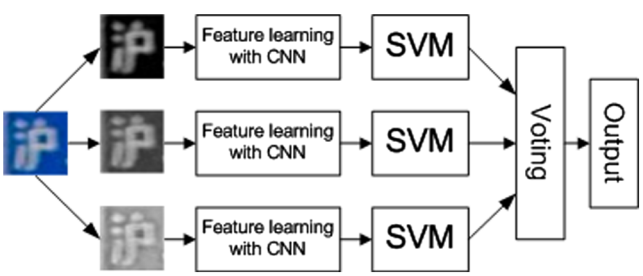

Fig. 5 Structure of the proposed classifier for Chinese character recognition.

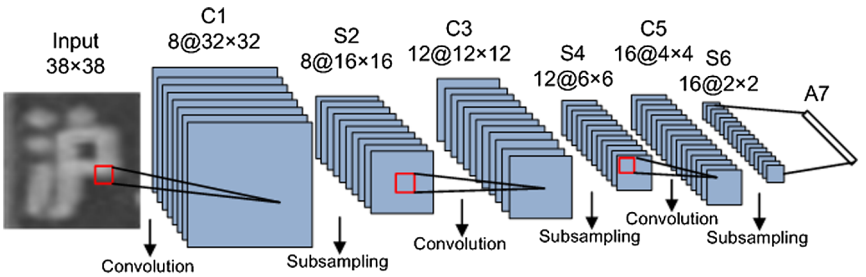

Fig. 6 Architecture of the three-convolutional-layer deep network.

changes. There are eight different 7×7 kernels in layer $C1$. To avoid convolution at the boundary, the size of every feature map is reduced to 32×32 units by skipping three pixels along the left, right, top, and bottom boundaries of the input image. In order to reduce the resolution of feature maps and also the sensitivity of the output to distortions, the subsampling operation is introduced to generate subsampling layers. There are eight subsampled feature maps at the first subsampling layer $S2$. Each neuron inside a subsampled feature map is connected to the 2×2 receptive field of the corresponding map in the previous layer $C1$ and weights for these four connections are all set as 0.25, which implies an average operation. Unlike the receptive fields in layer $C1$, the subsampling layer has nonoverlapped receptive fields and this results in a reduction of the resolution. In this way, each feature map at layer $S2$ corresponds to 16×16 neurons.

The second convolutional layer $C3$ has 12 different 5×5 kernels. Each map in layer $S2$ can generate 12 convolved feature maps, and this results in a total number of 96 since there are eight maps in layer $S2$. Every group of eight convolved feature maps generated from the same kernel can be merged into one map by an average operation. Consequently, layer $C3$ contains 12 feature maps. Similarly, to avoid convolution at the boundary, the size of every feature map in layer $C3$ is reduced to 12×12 units by skipping two pixels along the left, right, top, and bottom boundaries of the map in layer $S2$. Subsmapling maps of layer $C3$ leads to a reduction factor of 2 for both the horizontal and vertical directions. In such a case, each feature map in layer $S4$ corresponds to 6×6 neurons. There are 16 different 3×3 convolution kernels in layer $C5$. Convolving maps in layer $S4$ with these 16 kernels yield 192 convolved feature maps. Averaging every group of 12 maps generated from the same kernel results in 16 feature maps at layer $C5$. Due to the 3×3 convolution kernel, the size of the feature map at layer $C5$ is decreased to 4×4 neurons to skip the convolution at the boundaries. After subsampling layer $C5$, the third subsampled layer $S6$ includes 16 feature maps, where each one contains 2×2 neurons. The last layer $A7$ is a fully connected layer which contains 64 nodes; it is also the input layer of SVM.

Classical SVM can only handle two-class problems. In order to build a multiclass SVM, we construct k two-class SVMs with k indicating the number of classes, which means the output layer of the multiclass SVM has k nodes. The hidden layer consists of some Gaussian functions which are chosen as the nonlinear kernels to map the input space into a higher dimensional space. Positive samples of the i'th class are used to train the i'th two-class SVM, and the remaining two-class SVMs are trained with negative samples. The Gaussian function is chosen as the nonlinear kernel to map the input space into a higher dimensional space.

The training method of the proposed CNN is quite similar to the traditional one, where the standard backpropagation learning algorithm[48] is used. Given N training samples and M classes, the error function of the proposed CNN is

$$E = \frac{1}{2}\sum_{n=1}^{N}(\mathbf{t}_n - \mathbf{y}_n)^2, \tag{8}$$

where $\mathbf{t}_n$ is a vector with M dimensions representing the ground truth of the n'th sample and $\mathbf{y}_n$ is the output value of the CNN. For the n'th sample, its output at any layer l from the proposed CNN reads

$$\mathbf{y}_n^l = f\left(\sum \mathbf{y}_n^{l-1}\mathbf{k}^l + \mathbf{b}^l\right), \tag{9}$$

where $\mathbf{k}^l$ indicates the weight vector, $\mathbf{b}^l$ refers to the addition bias, and f means the sigmoid function.

The purpose of training is to minimize the error function by finding the proper value of weight vector $\mathbf{k}^l$. In this paper, we employ the gradient descent method to update the network weights. The updated expression of the weight vector is

$$\Delta\mathbf{k}^l = -\eta\frac{\partial E}{\partial \mathbf{k}^l}, \tag{10}$$

where η represents the learning rate.

3 Experimental Results

In this section, we present some experimental results to test the performance of our method. The type of video camera used in this paper is a Hikvision DS-2CD3T20D, manufactured by the Chinese company Hikvision. This camera has two mega pixels, its highest resolution is 1920×1080 and the frame rate is 30 fps. The camera is installed at the intersection of roads with a height of 6 m. It works 24 h with a light-emitting diode (LED) fill light. In the daytime, the illumination of images may vary depending on the weather conditions. However, since the LED fill light is always on, there is no dramatic illumination variation. In the night, due to the light reflection, license plates become brighter than the car body and they are easier to detect. When illumination gets lower, images can contain some noise. Experiments are tested on an Intel Core 2 Duo 2.2 GHz desktop computer with 4GB RAM. Currently, the development environment of our experiment is MATLAB® 2010a, although we are

Fig. 7 Top row: three car images segmented from the traffic video. Middle row: corresponding saliency maps for license plates. Bottom row: detected license plates.

planning to switch the platform to a C++ based one to achieve a better real-time performance.

Figure 7 shows some license plate detection results. The top row contains three cars segmented from the original traffic video and the middle row demonstrates the corresponding saliency maps generated using the modified visual attention model. The bottom row indicates the detected license plates.

In this paper, we use the recall and precision rates as the evaluation metric. The recall and precision rates R can be defined as $R = \mathrm{TP}/(\mathrm{TP}+\mathrm{FN})$ and $P = \mathrm{TP}/(\mathrm{TP}+\mathrm{FP})$, where TP means true positive, FN refers to false negative, and FP indicates false positive. We have used 835 separate car images in this experiment, Table 1 illustrates the recall and precision rates of the edge-based method and our approach. The edge-based method has a high recall rate of 98.1%, but our approach can achieve a slightly better result with a rate of 99.2%. For some images, more than one license plate region is detected; therefore, the precision rates are lower than the recall rates, and our method obtains a 1% increase when compared with the edge-based one. It is proven that the modified visual attention model can be used as an effective way for license plate detection and it outperforms the edge-based detection method.

We apply the popular projection method to segment the detected license plate region into seven blocks. For correctly detected license plates, the segmentation accuracy for Chinese characters, letters, and numbers, is 100%. Figure 8 shows three detected license plates and their corresponding segments. These segmented blocks are then normalized to a size of 38×38 pixels and directly used as inputs for classifiers.

In this paper, the first classifier is used to identify Chinese characters, numbers, and alphabet letters which can be recognized using the second classifier. The numbers of training samples for classifiers 1 and 2 are 930 and 1020, respectively, and the corresponding test sample numbers are 620 and 680. Convolutional neural networks with the same structure, as shown in Fig. 6, are employed to learn features in a hierarchical way for every color channel. Compared with the work in Ref. 42, the classification part is replaced by SVM and the convolutional layers are increased to three. Training samples are used to train the three-convolutional-layer networks and the feature weights are frozen to train corresponding SVMs.

The red channels of trained convolutional kernels for recognizing Chinese characters, numbers, and alphabet letters are shown in Figs. 9 and 10, respectively. In each figure, the top row illustrates eight kernels at the first convolution layer. Even though we have used 12 and 16 convolutional kernels for the second and third convolution layers, only eight from each of them are selected and demonstrated.

Convolving the kernels in Fig. 9 with the left most segment of the license plate, i.e., the Chinese character which

Table 1 Recall rates for license plate detection.

Methods	Recall rate (%)	Precision rate (%)
Edge based	98.1	97.9
Our method	99.2	98.9

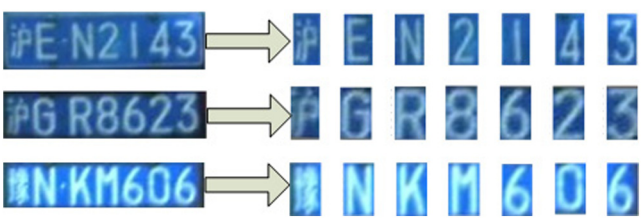

Fig. 8 Three license plates and the corresponding segmented blocks.

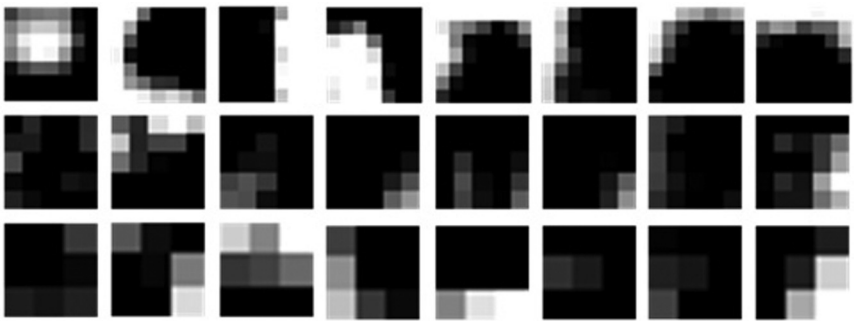

Fig. 9 Convolutional kernels for Chinese character recognition. Top row: 8 kernels for the first convolutional layer. Middle and bottom rows: selected kernels for the second and third layers.

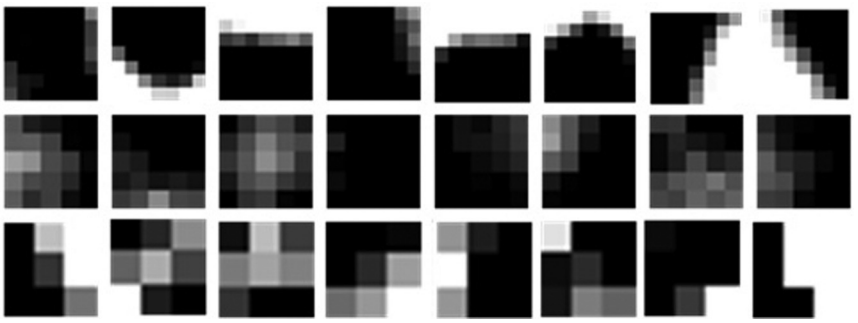

Fig. 10 Convolutional kernels for numbers and alphabet letter recognition. Top row: 8 kernels for the first convolution layer. Middle and bottom rows: selected kernels for the second and third layers.

Fig. 11 An example of Chinese character convolution results using the deep convolutional neural network. Top row: the first convolutional layer with eight feature maps. Middle row: 12 feature maps of the second convolutional layer. Bottom row: the third convolutional layer, which includes 16 feature maps.

depicts the region to which the car belongs, can yield corresponding feature maps as shown in Fig. 11. The top row contains eight feature maps of the first convolution layer, and 12 and 16 feature maps for the second and third convolutional layers are displayed in the middle and bottom rows. Figure 12 illustrates the corresponding feature maps for a given input letter "A."

It takes 20 ms to recognize a single segment and the time required to identify a Chinese license plate is 140 ms. To check the performance of our approach, we compare it with two different methods. The first one is a frequently used SVM-based recognition approach. The feature vectors of inputs are extracted using the scale invariant feature transform (SIFT) descriptor,[49] and the classification is completed

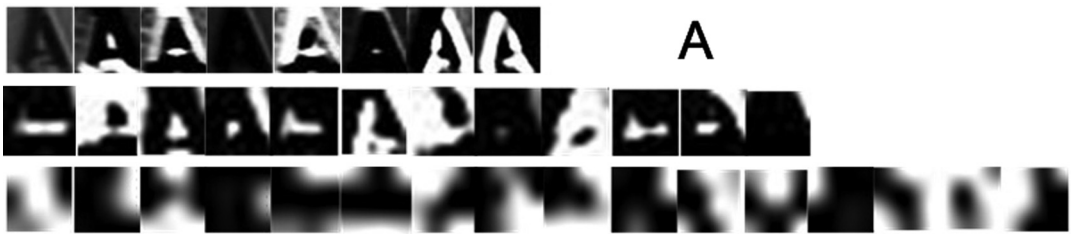

Fig. 12 An example of letter convolution results using the deep convolutional neural network. Top row: the first convolutional layer with eight feature maps. Middle row: 12 feature maps of the second convolutional layer. Bottom row: the third convolutional layer, which includes 16 feature maps.

with SVM. The second method is the traditional CNN proposed in Ref. 42, where the number of convolution layers is two and the classification part contains two full connection layers and one Gaussian connection layer. Table 2 demonstrates the average comparison results of recall rates.

In this table, "SIFT + SVM" represents the SVM-based approach, "CNN2" refers to the regular CNN proposed in Ref. 42, "Multichannel + CNN3 + SVM" indicates our approach. In order to analyze the contribution of different stages in the proposed method, we add the included stages of "CNN3" and "CNN3 + SVM," where "CNN3" means the three-convolutional-layer network with the same classification part as "CNN2" and "CNN3 + SVM" indicates the neural network with three convolutional layers with its classification replaced by SVM. "Multichannel + CNN3 + SVM" represents the proposed classifier which integrates three color channels followed by a voting procedure where every channel consists of a CNN with three convolutional layers and its classification is completed by SVM.

In contrast to Chinese characters, numbers and letters contain less structural information and are easier to recognize; hence, recall rates are higher than that of Chinese characters for all methods. For the "SIFT + SVM" method, recall rates are 91.1% and 93.1% for Chinese characters, numbers, and letters, respectively. Because "CNN2" can automatically learn features, the recall rate increases 3.9% on the average. Compared with "CNN2," "CNN3" has a deeper architecture, which results in an average growth of 1.85%. By replacing the classification part with SVM, "CNN3 + SVM" gains an average recall rate of 98.5% with an increase of 0.6% when compared with "CNN3." This indicates that SVM can only contribute a bit to the performance improvement. By learning features from three color channels followed by a voting mechanism, the method of "Multichannel + CNN3 + SVM" can obtain an average rise of 0.25%.

Table 3 illustrates precision rates for all methods. Since there are more false-positive results than false-negative results, the precision rates are, in general, lower than the recall rates. Compared with "SIFT + SVM," "CNN2" achieves an average growth of 3.85%; the precision rate of "CNN3" is, on average, 2.05% higher than that of "CNN2"; on the basis of "CNN3," "CNN3+SVM" gains a 0.55% improvement and "Multichannel + CNN3 + SVM" again obtains a 0.2% increase. Results in Tables 2 and 3 indicate that a deep neural architecture for feature learning is the most important contribution to the recognition improvement.

Table 2 Recall rate comparison.

Methods	Chinese characters (%)	Numbers and letters (%)
SIFT + SVM	91.1	93.1
CNN2	95.2	96.8
CNN3	97.3	98.4
CNN3 + SVM	98.0	98.9
Multichannel + CNN3 + SVM	98.3	99.1

Table 3 Precision rate comparison.

Methods	Chinese characters (%)	Numbers and letters (%)
SIFT + SVM	90.3	92.4
CNN2	94.3	96.1
CNN3	96.7	97.8
CNN3 + SVM	97.2	98.4
Multichannel + CNN3 + SVM	97.4	98.6

As for the classification part, SVM contributes only 0.6% and 0.55% to the increase of the recall and precision rates; multichannel processing and the voting mechanism result in a slight growth in performance.

The regular CNN takes raw segments of the license plate as direct inputs. Due to the deep network structure, it can learn a hierarchy of features which enable better descriptions of the inputs. Therefore, it performs much better than the SVM-based one. The traditional CNN can only handle gray inputs; however, we design two new classifiers to handle color inputs by cascading CNNs and SVMs for processing three color channels. In addition, CNNs used in the proposed classifiers have deeper structures with three convolutional layers. Experimental results prove that our method outperforms both the SVM-based one and the regular CNN method.

Even though the proposed method has a good performance, there are still some unsuccessful cases needing improvement. For every segmented block, before the normalization, the original width and height are around 19 and 40 pixels. The resolution of each segment is relatively low, which makes it difficult to correctly recognize Chinese characters with a lot of structures. Due to the noise, low resolution segments of the digital number 2 and letter "Z" can hardly be distinguished. In addition, car images are separated from traffic videos using motion information which can result in some blurred license plates that may yield incorrect recognition results.

In order to check the robustness of the proposed method in the environment of illumination variation and noise contamination, as shown in Fig. 13, we manually add illumination change and some noise to the test samples. The top row shows Chinese character segments with gamma parameters of 0.3, 0.7, and 1.5. The corresponding segments added with salt and pepper noise percentages of 5, 15, and 25 are displayed in the bottom row.

Recall rates under the condition of illumination change are displayed in Fig. 14. The blue and red curves represent recall rates for Chinese characters, numbers, and letters, respectively. It is shown that correct recognition results are all above 90% when the gamma parameters of image segments are within the range of 0.4 and 2.5, which indicates that the proposed method can work robustly in an environment of illumination variation.

Figure 15 illustrates the change of recall rates with respect to the percentages of salt and pepper noise. Because Chinese

Fig. 13 Chinese character segments with different illuminations and noise contamination. Top row: from left to right are segments with gamma parameters of 0.3, 0.7 and 1.5. Bottom row: from left to right are salt and pepper noise contaminated images with noise percentages of 5, 15, and 25.

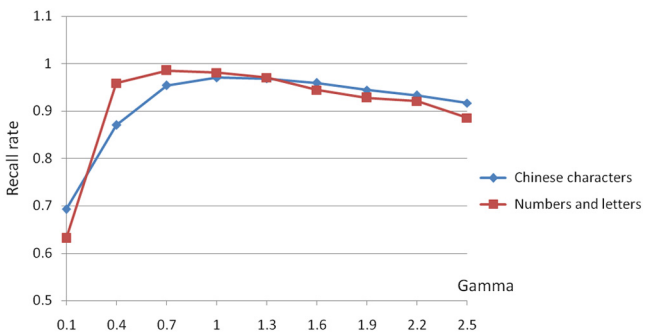

Fig. 14 The change of recall rates with respect to the variation of gamma values.

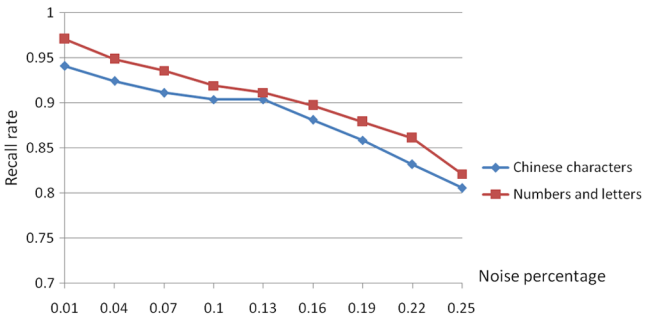

Fig. 15 The change of recall rates with respect to percentages of salt and pepper noise.

characters have more detailed structure information, recall rates of numbers and letters under the noise contamination condition are slightly better. However, even though the percentage of salt and pepper noise increases to 25%, all recall rates are still higher than 80%.

4 Conclusions

In this paper, we present a new method to recognize Chinese car license plates in traffic videos. Based on multiple frames in the temporal domain, car images are separated from traffic videos by computing the motion information in regions of interest. Once car images are preprocessed to enhance contrast and remove noise, we use a modified visual attention model to detect the license plate. The color BY, as the combination of blue and yellow, is used to generate a feature map which can support accurate detection in the context of a Chinese license plate. In addition, the saliency map is finely tuned according to the prior information of the license plate. In order to segment the detected license plate into seven blocks, we employ the fast and accurate projection method. Two new classifiers are proposed to recognize Chinese characters, numbers, and alphabet letters. Each classifier integrates the CNN and SVM into a single framework, and three color channels are simultaneously processed for yielding the final result via a majority voting process. Demonstrated results prove that the proposed method has high recall and precision rates, and works robustly under the environment of illumination change and noise contamination.

Acknowledgments

This work has been supported by National Natural Science Foundation of China (Nos. 61103071, 61103072, 61272271, 61472284), Natural Science Foundation of Shanghai, China (No. 12ZR1434000, 13ZR1443100) and Research Fund for the Doctoral Program of Higher Education of China (No. 20110072120065).

References

1. C. Anagnostopoulos et al., "A license plate recognition algorithm for intelligent transportation system applications," *IEEE Trans. Intell. Transp. Syst.* **7**(3), 377–392 (2006).
2. J. Jiao, Q. Ye, and Q. Huang, "A configurable method for multi-style license plate recognition," *Pattern Recognit.* **42**, 358–369 (2009).
3. C. Anagnostopoulos et al., "License plate recognition from still images and video sequences: a survey," *IEEE Trans. Intell. Transp. Syst.* **9**(3), 377–391 (2008).
4. H. Sheng et al., "Real-time anti-interference location of vehicle license plates using high-definition video," *IEEE Intell. Transp. Syst. Mag.* **1**(4), 17–23 (2009).
5. Y. Qiu, M. Sun, and W. Zhou, "License plate extraction based on vertical edge detection and mathematical morphology," in *Proc. Int. Conf. on Computer Intelligent Software Engineering*, pp. 1–5 (2009).
6. G. Hsu, J. Chen, and Y. Chung, "Application-oriented license plate recognition," *IEEE Trans. Veh. Technol.* **62**(2), 552–561 (2013).
7. A. M. Al-Ghaili et al., "Vertical edge based car license plate detection method," *IEEE Trans. Veh. Technol.* **62**(1), 26–38 (2013).
8. M. H. T. Brugge et al., "License plate recognition using DTCNNs," in *Proc. 5th IEEE Int. Workshop on Cellular Neural Networks and Their Applications*, pp. 212–217 (1998).
9. C. N. E. Anagnostopoulos et al., "A license plate recognition algorithm for intelligent transportation system applications," *IEEE Trans. Intell. Trans. Syst.* **7**(3), 377–392 (2006).
10. W. G. Zhu, G. J. Hou, and X. Jia, "A study of locating vehicle license plate based on color feature and mathematical morphology," in *Proc. 6th Int. Conf. on Signal Process*, pp. 748–751 (2002).
11. S. Chang et al., "Automatic license plate recognition," *IEEE Trans. Intell. Transp. Syst.* **5**(1), 42–53 (2004).
12. X. Shi, W. Zhao, and Y. Shen, "Automatic license plate recognition system based on color image processing," *Lec. Notes Comput. Sci.* **3483**, 1159–1168 (2005).
13. S. Gendy, C. L. Smith, and S. Lachowicz, "Automatic car registration plate recognition using fast Hough transform," in *Proc. 31st Ann. Int. Carnahan Conf. on Security Technology*, pp. 209–218 (1997).
14. Y. Yanamura et al., "Extraction and tracking of the license plate using Hough transform and voted block matching," in *Proc. IEEE Intell. Vehicles Symp.*, pp. 243–246 (2003).

15. W. Zhou, H. Li, and Y. Lu, "Principal visual word discovery for automatic license plate detection," *IEEE Trans. Image Process.* **21**(9), 4269–4279 (2012).
16. S. K. Kim, D. W. Kim, and H. J. Kim, "A recognition of vehicle license plate using a genetic algorithm based segmentation," in *Proc. Int. Conf. on Image Processing*, pp. 661–664 (1996).
17. H. Zhang et al., "Learning-based license plate detection using global and local features," in *Proc. 18th Int. Conf. on Pattern Recognition*, pp. 1102–1105 (2006).
18. M. Zheng and Q. Liu, "Application of LVQ neural network to car license plate recognition," in *Proc. Int. Conf. on Intelligent Systems and Knowledge Engineering*, pp. 287–290 (2010).
19. Y. Chen et al., "The application of a convolution neural network on face and license plate detection," in *Proc. Int. Conf. on Pattern Recognition*, pp. 552–555 (2006).
20. Z. Zhao, S. Yang, and X. Ma, "Chinese license plate recognition using a convolutional neural network," in *Proc. IEEE Pacifc-Asia Workshop on Computational Intelligence and Industrial Application*, pp. 27–30 (2008).
21. X. Chen et al., "Vehicle detection in satellite images by hybrid deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.* **11**(10), 1797–1801 (2014).
22. J. R. Anderson, Ed., *Cognitive Psychology and Its Implications*, 6th ed., Worth Publishers, New York (2005).
23. B. A. Olshausen, C. H. Anderson, and D. C. V. Essen, "A neuro biological model of visual attention and invariant pattern recognition based on dynamic routing of information," *J. Neuroscience* **13**(11), 4700–4719 (1993).
24. J. K. Tsotsos et al., "Modelling visual attention via selective tuning," *Artif. Intell.* **78**(1–2), 507–545 (1995).
25. L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(11), 1254–1259 (1998).
26. E. Niebur and C. Koch, *Computational Architectures for Attention*, pp. 163–186, MIT Press, Cambridge, Massachusetts (1998).
27. S. B. Choi et al., "Biologically motivated vergence control system using human-like selective attention model," *Neuro Comput.* **69**(4–6), 537–558 (2006).
28. S. Ban, I. Lee, and M. Lee, "Dynamic visual selective attention model," *Neuro Comput.* **71**, 853–856 (2008).
29. Q. Wu et al., "A visual attention model based on hierarchical spiking neural networks," *Neuro Comput.* **116**, 3–12 (2013).
30. T. D. Duan et al., "Building an automatic vehicle license-plate recognition system," in *Proc. Int. Conf. on Computer Science*, pp. 59–63 (2005).
31. Z. Qin et al., "Method of license plate location based on corner feature," in *Proc. World Congr. Intelligent Control Automation*, pp. 8645–8649 (2006).
32. Y. Wen et al., "An algorithm for license plate recognition applied to intelligent transportation system," *IEEE Trans. Intell. Transp. Syst.* **12**(3), 830–845 (2011).
33. S. Nomura et al., "A new method for degraded color image binarization based on adaptive lightning on grayscale versions," *IEICE Trans. Inf. Syst.* **E87-D**(4), 1012–1020 (2004).
34. S. Nomura et al., "A novel adaptive morphological approach for degraded character image segmentation," *Pattern Recognit.* **38**(11), 1961–1975 (2005).
35. P. Comelli et al., "Optical recognition of motor vehicle license plates," *IEEE Trans. Veh. Technol.* **44**(4), 790–799 (1995).
36. X. Lu, X. Ling, and W. Huang, "Vehicle license plate character recognition," in *Proc. Int. Conf. on Neural Network Signal Processing*, pp. 1066–1069 (2003).
37. M. Kseneman and D. Gleich, "License plate recognition using feedforward neural networks," *J. Microelectron. Electron. Compon. Mater.* **41**(3), 212–217 (2011).
38. Y. Liu et al., "Vehicle-license-plate recognition based on neural networks," in *Proc. IEEE Int. Conf. on Information and Automation*, pp. 363–366 (2011).
39. M. Ghazal and H. Hajjdiab, "License plate automatic detection and recognition using level sets and neural networks," in *Proc. Int. Conf. on Communications, Signal Processing, and Their Applications*, pp. 1–5 (2013).
40. Y. Amit, D. Geman, and X. Fan, "A coarse-to-fine strategy for multiclass shape detection," *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(12), 1606–1621 (2004).
41. C. Arth, F. Limberger, and H. Bischof, "Real-time license plate recognition on an embedded platform," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1–8 (2007).
42. Y. LeCun, L. B. Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE* **86**(11), 2278–2324 (1998).
43. M. Norouzi, M. Ranjbar, and G. Mori, "Stacks of convolutional restricted boltzmann machines for shift-invariant feature learning," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 2735–2742 (2009).
44. S. Ji et al., "3-D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(1), 221–231 (2013).
45. H. Cecotti and A. Graser, "Convolutional neural networks for P300 detection with application to brain-computer interfaces," *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(3), 433–445 (2011).
46. H. Lee et al., "Unsupervised feature learning for audio classification using convolutional deep belief networks," in *Proc. Advances in Neural Information Processing Systems*, pp. 1096–1104 (2009).
47. N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst. Man Cybernet.* **9**(1), 62–66 (1979).
48. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature* **323** 533–536 (1986).
49. D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th Int. Conf. on Computer Vision*, pp. 1150–1157 (1999).

Di Zang is an associate professor with the Department of Computer Science and Technology of Tongji University of China. Her current research interests include machine learning, pattern recognition, and intelligent transportation.

Zhenliang Chai received his BS and MS degrees from the Department of Computer Science and Technology of Tongji University of China. His research areas include pattern recognition, machine learning and intelligent transportation. He is now an engineer at Baidu of China.

Junqi Zhang is an associate professor with the Department of Computer Science and Technology of Tongji University of China. His research interest includes intelligent computing, learning automata, machine learning, and sequential optimization.

Dongdong Zhang is an associate professor with the Department of Computer Science and Technology of Tongji University of China. Her research interests include high-definition video coding, perceptual video coding, scalable video coding, three-dimensional (3-D) video coding and rendering, and so on.

Jiujun Cheng is presently an associate professor of Tongji University, Shanghai, China. He has over 40 publications, including conference and journal papers. His research interests include area of mobile computing and social network with a focus on mobile/Internet interworking, and Internet of vehicles.