Multi-Dimensional Traffic Congestion Detection Based on Fusion of Visual Features and Convolutional Neural Network

Xiao Ke, Lingfeng Shi, Wenzhong Guo, *Member, IEEE*, and Dewang Chen, *Senior Member, IEEE*

***Abstract*— In intelligent transportation systems, there are many tasks that rely on the detection of road congestion, such as traffic signal scheduling and traffic accident detection. As traditional methods for traffic congestion detection are difficult to use, expensive, and may cause damage to the road surface, this paper presents a method for road congestion detection that is based on multidimensional visual features and a convolutional neural network (CNN). This method first detects the density of foreground objects by using a gray-level co-occurrence matrix; second, the speed of moving objects is detected by using the Lucas–Kanade optical flow with pyramid implementation. Third, a Gaussian mixture model is used to model the background, and the CNN is then used to accurately detect the final foreground from the candidate foregrounds. Finally, the proposed method performs road congestion detection in terms of a multidimensional feature space, including traffic density, traffic velocity, road occupancy, and traffic flow. Furthermore, we propose an information entropy method using a histogram of optical flow to enhance the accuracy and reliability of road congestion detection. Simulation results via quantitative and qualitative assessment indicate that the proposed method is able to significantly outperform the state-of-the-art road-traffic congestion detection methods due to the fusion of multidimensional features using the CNN.**

***Index Terms*— Traffic congestion detection, convolutional neural network (CNN), gray-level co-occurrence matrix (GLCM), optical flow.**

Manuscript received May 19, 2017; revised December 25, 2017 and May 24, 2018; accepted July 12, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61502105, Grant 61672159, and Grant 61672158, in part by the Technology Guidance Project of Fujian Province under Grant 2017H0015, in part by the Technology Innovation Platform Project of Fujian Province under Grant 2014H2005, in part by Fujian Natural Science Funds for the Distinguished Young Scholar under Grant 2015J06014, in part by the University Production Project of Fujian Province under Grant 2017H6008, in part by the Fujian Collaborative Innovation Center for Big Data Application in Governments, and in part by the Fujian Engineering Research Center of Big Data Analysis and Processing. The Associate Editor for this paper was K. Wang. *(Corresponding author: Wenzhong Guo.)*

X. Ke and W. Guo are with the College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350116, China, also with the Fujian Provincial Key Laboratory of Networking Computing and Intelligent Information Processing, Fuzhou University, Fuzhou 350116, China, and also with the Key Laboratory of Spatial Data Mining & Information Sharing, Ministry of Education, Fuzhou 350003, China (e-mail: guowenzhong@fzu.edu.cn).

L. Shi and D. Chen are with the College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350116, China, and also with the Fujian Provincial Key Laboratory of Networking Computing and Intelligent Information Processing, Fuzhou University, Fuzhou 350116, China.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2018.2864612

I. Introduction

With the acceleration of urbanization, traffic problems are aggravated, resulting in economic losses and paralysis of urban functions [1]. Moreover, vehicle energy consumption and environmental pollution caused by road congestion are worsening. Therefore, many papers focus on research on intelligent transportation systems (ITS) [2]–[5], and research on automatic road congestion detection has become the focus of attention. Marfia and Roccetti [6] calculate two ratios to detect congestion, one with respect to the number of vehicles in congestion and leaving congestion, and the other with respect to the number of vehicles entering congestion according to traversal times collected during an observation period for a given road segment. In [7], five traffic rerouting strategies are designed to be incorporated in a traffic guidance system. To operate this system, the traffic congestion prediction is executed in the second phases based on road speed and number of vehicles. Cao *et al.* [8] proposed a unified traffic management framework for vehicle rerouting and traffic light control. To performing vehicle rerouting or traffic light control, pheromone-based traffic condition prediction is implemented in the proposed framework. Terroso-Saenz *et al.* [9] detect traffic situations by stating that Complex Event Processing (CEP) is suitable for processing beacon messages from a vehicular ad hoc network, and they proposed a CEP-based event-driven architecture for distributed traffic information systems based on the continuous exchange of information between vehicles.

The above methods, however, all rely on some prior knowledge of the roads, such as the length of the road, number of lanes, traffic light cycle, and real-time vehicle information. Furthermore, in the process of collecting such information in traditional ITSs, hardware facilities are needed, such as ground sense coil [10], GPS [11], inductive loop detector [12], and so on. Note that the construction of the ground sense coil may, however, damage the road surface, and its construction is complicated and difficult to maintain [1]. GPS-based ITSs depend upon the number of vehicles that use GPS, meaning the detection accuracy of traffic status is greatly reduced when the number of vehicles using GPS significantly decreases [11]. Therefore, many techniques for ITSs based on surveillance videos have been proposed with the development of image processing technology. These techniques have become popular due to the fact that they not only prevent damage to road

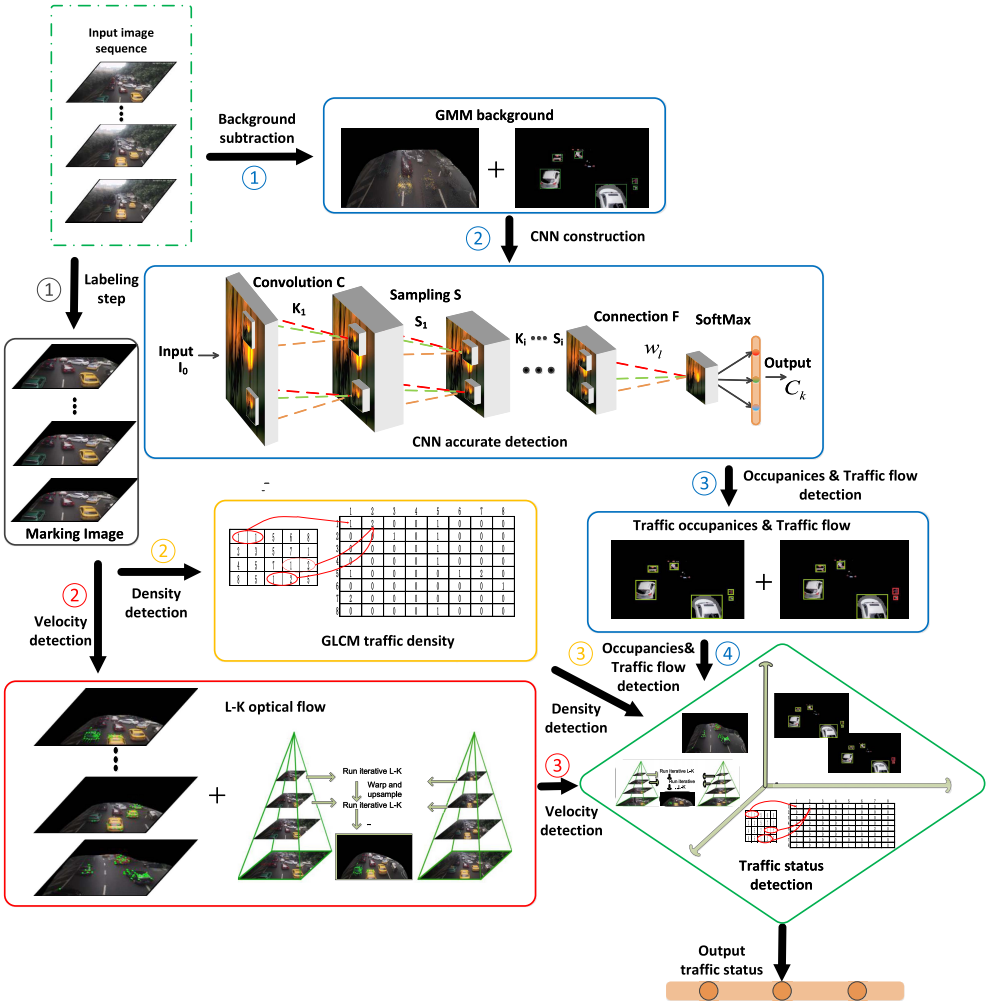

Fig. 1. Procedure for the proposed approach. The procedures at the same level of processing are represented by same numeric label. The processes in the same flow path are denoted by the same label color.

surfaces, but reflect and deliver large-scale traffic information in real time.

Hsieh *et al.* [13] achieves the classification of vehicles based on surveillance videos by introducing a linearity feature, which uses a line-based algorithm for shadow removal, and a Kalman filter to detect foreground. In [14], a moving object is extracted by using adaptive background estimation and Gaussian shadow elimination, which makes use of virtual detection instead of the traditional inductive loop detectors. However, [14] does not implement recognition and classification of the moving objects in the foreground. Song *et al.* [15] proposed an adaptive method for the background model, which implements background subtraction and vehicle counting based on a block. Nonetheless, this method has difficulty dealing with traffic congestion conditions or other complex traffic circumstances. Reference [16] is similar to [14] in that it also makes use of virtual detection for vehicle counting, and then achieves real-time traffic monitoring via a background segmentation algorithm based on approximated median filter for motion detection. The major downside of [16] resides in the low detection accuracy resulting from the shadow of foreground objects.

In this paper, we propose a novel traffic-congestion detection model for traffic-surveillance automatic video analysis by carefully considering the practical factors as shown in Fig. 1, that closely relate to changes in traffic conditions.

(1) A GMM-based CNN foreground detection for complex real traffic circumstance. In order to obtain precise foreground motion, we build a CNN to accurately detect foreground objects based on GMM to further objects detection at the GMM stage. By accurate foreground estimation with a combination of GMM and CNN, the model can thereby tackle a complex traffic circumstance with intricate background and multiple moving targets.

(2) An efficient congestion detection using multidimensional visual features. Instead of only describing a traffic jam in terms of a single feature dimension as most traditional methods do, the proposed method comprehensively describes the congestion status from a multidimensional feature space, including traffic density, traffic speed, traffic occupancy, and traffic flow. Through the complement and interaction of these multidimensional visual features from the dynamic and static perspectives, we can better portray the state of traffic congestion.

(3) Entropy-based HOF for camera displacements. We propose an information entropy-based HOF to handle the error of optical flow calculations caused by camera displacements in traffic-speed estimation. The information entropy-based HOF not only effectively eliminates the optical flow error caused by sudden displacements of the camera, but detects pixel matching errors between two sequential frames.

With better accuracy in foreground-object detection, and comprehensive visual feature analysis by the information entropy-based HOF, we can more effectively detect traffic congestion via the proposed multidimensional detection model for traffic surveillance videos. Experimental results obtained by visual and quantitative evaluation demonstrate that the proposed model outperforms the existing state-of-the-art road traffic congestion detection methods on tested traffic surveillance videos.

The following sections of this paper are organized as follows. An introduction to the GMM background model and the CNN constructed for the accurate detection of foreground targets is presented in Section II; The integration of multidimensional visual features with the CNN is discussed in Section III, including road occupancy, traffic flow, road density, and traffic velocity. In Section IV, we investigate how to detect congestion status using multidimensional features. Section V discusses the experimental results compared with other state-of-the-art algorithms. Finally, our conclusions are drawn in Section VI.

II. Target Foreground Detection Based on CNN

In this paper, our first goal is to achieve accurate detection of moving objects via the fusion of a GMM detector and a CNN classifier. The subsequent congestion detection can then be obtained based on foreground object detection. To detect the objects in a dynamic video that includes multiple moving objects, especially traffic surveillance videos, the motion information in the dynamic video can be used to facilitate the efficient detection of moving objects.

To this end, in traffic surveillance, the foreground detection method that combines background modeling by CNN and GMM can make full use of motion information between frames in the dynamic video.

A. The Underlying Background Estimation With GMM

To obtain the road occupancy and traffic flow, a background model is needed so that the foreground objects can be extracted from the image. The scenes of roads which are filmed by cameras for road traffic detection generally have the following four characteristics: 1. All-weather shooting with gradual changed light; 2. Camera placed in the outdoors and is vulnerable to outdoor environment and weather; 3. The camera is generally placed high, some cameras are covered by leaves; 4. The branches and leaves obvious shakes in some scenes. Consequently, a background modeling method that can be applied to such complex background scenario is required.

The Vibe [17] is a non-parametric foreground detection algorithm which adopts the random background updating strategy. Vibe has the advantages of fast background modeling and small computation, but there are some problems such as the inability to eliminate the shadow, the incomplete target in the detection and the existence of ghosts. GMM [18]–[20] not only describes the multimodal status of the pixels but eliminates repeatedly intricate disturbances from moving background such as illumination variation, leaf shaking, and noise effect, GMM accordingly model complex background accurately and apply to a variety of complex background and widely used in various scenes successfully. GMM is superior to ViBe algorithm in the accuracy of change detection and small moving target detection.

Taking into account the characteristics of the GMM method and traffic video, this paper adopts the GMM as the background modeling method.

If traffic is congested for a long period of time, it is found that the scene, the height and the angle of the surveillance camera will affect the degradation of GMM to a certain extent. Thus the estimation of road occupancy and traffic flow by the GMM will be affected. Assuming that road occupancy and traffic flow is incorrectly detected as zero because of the degradation of the GMM caused by long-duration congestion. However, density detection will still work under such a congestion situation to detect congestion via our proposed multidimensional model, which can be configured according to the specific requirements of the traffic department include different scenarios, different monitoring cameras configuration and the congestion detection area.

In short, one attribute of the proposed multidimensional congestion detection model is that the accuracy of congestion detection will not greatly decline even if the GMM is badly degraded when traffic is congested for a long period of time, and the model based on multidimensional visual features maintains good congestion detection results nonetheless.

B. CNN for Precise Target Detection

In order to further improve the detection accuracy of road traffic foreground targets, this paper proposes a method for more accurate foreground detection by constructing a CNN classifier basd on Alexnet structure [21], [22]. By training the CNN, the input image patterns (i.e., the input candidate foreground objects) are sorted by scoring with the trained neural network. Comparing the score of each category with the highest score will then determine which class the foreground object belongs to.

In this study, the objects are generally divided into three classes: vehicles, scooters, and pedestrians or outliers. According to the 2D imaging property, the characteristics of a scooter is very similar to that of a pedestrian in a captured frame. Therefore, this study classifies scooters and pedestrians into the same category. As a result, the total objects category C can be defined as follows:

$$\begin{aligned} C_1 &= \{d|d : \text{vehicles}\} \\ C_2 &= \{d|d : \text{pedestrians or scooters}\} \\ C_3 &= \{\sim C_1 \cap \sim C_2\}. \end{aligned} \tag{1}$$

The CNN constructed in this study is composed of 11 core network layers in total, including 5 convolutional layers,

TABLE I
INFORMATION OF CNN FOR 11 CORE NETWORK LAYERS

	C1	P2	C3	P4	C5	C6	C7	P8	FC9	FC10	FC11
The number of kernels	96	—	256	—	384	384	256	—	—	—	—
Kernel size & sampling window size	11×11	3×3	5×5	3×3	3×3	3×3	3×3	3×3	—	—	—
Step length	4	2	1	2	1	1	1	2	—	—	—
The number of feature maps to process	3	96	96	256	256	384	384	256	—	—	—
The number of neurons	—	—	—	—	—	—	—	—	4096	4096	1000

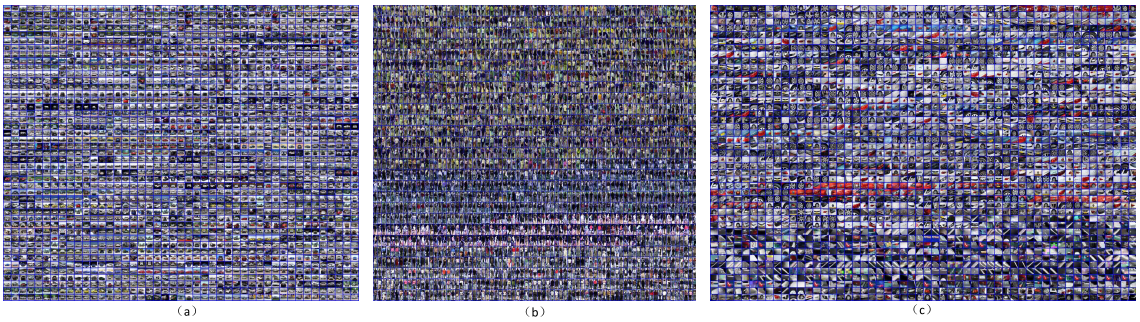

Fig. 2. The collected training data from website of three categories for CNN learning. (a) 2025 image for vehicles class. (b) 2025 image for pedestrians & scooters class. (c) 2045 image for error detection.

3 pooling layers, and 3 fully-connected layers. In addition, it also contains an input layer and a SoftMax layer. The information of CNN is given in Table I. In Table I, C denotes the convolutional layer, P is the pooling layer, and FC is the fully connected layer.

The CNN-based foreground object detection method in this paper consists of two phases. Firstly, we use the collected training data from China Telecommunications to learn a CNN classifier, including 2025 images of vehicle class, 2025 images of pedestrians & scooters class, and 2045 images for error detection, These training data are shown in Fig. 2. Secondly, by training the CNN, the foreground objects inside bounding boxes are sorted by scoring with the trained neural network. We select the category with the highest score as the attribution class via SoftMax layer in detection phase. We use 3-dimensional binary vector to represent its attribution classes. "1" means belonging, and "0" means that it does not belong. The learning rate and learning weight decay are fixed and used in the whole learning phase of algorithm 1. After the completion of the learning phase, the detection model will be obtained, and the two parameters of learning rate and learning weight decay need not be used in the detection phase. The procedure for CNN training and classification algorithm is given in Algorithm 1.

In doing so, we can more accurately detect the foreground by using the CNN, so we get the final foreground set $D = \{d_1, d_2, \ldots, d_m\}$.

III. MULTIDIMENSIONAL VISUAL FEATURES FOR TRAFFIC DETECTION

In Section II, we discussed the modeling process for the background by the GMM, and the accurate detection of the foreground by the CNN, which greatly improves the accuracy of target detection. Finally, we get the foreground object set D.

Due to the drawback that a single-dimension visual feature approach cannot be applied well in road status detection, a method that integrates CNN and multidimensional visual features is proposed to accurately describe the road traffic status. Specifically, suppose a vehicle is driving slowly or even stops while a road is under light traffic conditions, the traffic speed is not much more than that it would be under congested conditions. In order to describe the road status, the proposed method makes use of motion information and pixel-wise methods based on dynamic and static features.

In general, we consider that the probability of road congestion is greater under conditions of higher traffic density, traffic flow, road occupancy, and slower traffic velocity. In the following part, the connection between traffic congestion and each visual feature is explained.

A. Traffic Occupancy Detection

After acquiring foreground set D in the CNN step, the feature extraction process for road occupancy and traffic flow can be obtained simultaneously.

In the traffic status detection process, we note that the road occupancy of vehicles or pedestrians is significantly increased when the road is congested. As a result, we take account of traffic occupancy as one of the indicators of traffic status. In this study, road traffic occupancy refers to the proportional occupancy by road vehicles on the specified area of a road over a specified period of time.

There are two basic methods for calculating traffic occupancy [23]. One is based on pixel counting, and the

ALGORITHM 1 Foreground Objects Detection by using CNN

Input: Testing d, training data I, labels of training data c_i.
Output: The detected label C of d.
Initialization:
K : convolutional kernel between the neurons.
W : weights of full-connection layers.
b : bias between the neurons of layers.
Set number of CNN layers $M \leftarrow 13$.
Set number of iterations $T \leftarrow 80000$.
Set base learning rate $\leftarrow 0.001$.
Set learning weight decay $\leftarrow 0.0005$.
Set momentum $\leftarrow 0.9$.
1. Learning phase
// Transform labels into binary vector of 3 dimensions. $\{c_1, c_2, c_3\} = Y_3 \in \{0, 1\}^3$;
for $j = 1 : T$ **do**
  for $i = 1 : M$ **do**
    if $i ==$ CONVOLUTION **then**
      $I^{(i+1)} = f(K^{i+1} \otimes I^{(i)} + b^{(i+1)})$;
    else if $i ==$ POOLING **then**
      $I^{(i+1)} = pool(I^{(i)})$;
    else if $i ==$ FULL-CONNECTION **then**
      $S = f(W_i \cdot I^{(i-1)} + b^{(i)})$;
    else
      $Score = \frac{e^S}{\sum_{\forall S} e^S}$;
      // Calculate loss and update weights by BP algorithm.
      $L = -\log(Score)$;
    end if
  end for
  $T = \arg\min L$; //Minimize the loss.
end for
2. Detection phase
for $i = 1 : M$ **do**
  if $i ==$ CONVOLUTION **then**
    $d^{(i+1)} = f(K^{i+1} \otimes d^{(i)} + b^{(i+1)})$;
  else if $i ==$ POOLING **then**
    $d^{(i+1)} = pool(d^{(i)})$;
  else if $i ==$ FULL-CONNECTION **then**
    $S = f(W_i \cdot d^{(i-1)} + b^{(i)})$;
  else
    $Score = \frac{e^S}{\sum_{\forall S} e^S}$;
    $C = \arg \max_{k=1,2,3} [Score_k]$;
  end if
end for

other is based on area calculations. The former resorts to calculating the ratio of the number of pixels in the foreground objects to the number of pixels in the road background without foreground targets. When the road background is fixed, the count of foreground pixels represents the extent to which the foreground objects occupy the road. Specifically, the road occupancy ratio will increase with an increasing number of foreground objects, which means that the probability of road congestion is greater than before.

However, there is a hypothetical problem in calculating occupancy based on pixels. The assumption is that each of the pixels detected in the foreground represents a foreground object. In fact, this assumption is highly dependent on the integrity of foreground detection. In practice, pixels of foreground objects whose intensity are similar to background pixels are sometimes detected as background pixels, which is why holes often appear in the detected foreground.

Furthermore, the objects of interest (i.e., both C_1 and C_2 categories) in road traffic can be simply abstracted geometrically into a rectangle, including vehicles, scooters and pedestrians. As a consequence, an area-based method for road occupancy is more feasible than the pixel-based method we have discussed. An area-based method calculates the area of the Minimum Enclosing Rectangle (MER) of the connected areas of each foreground object. When the road background is fixed, the area of the MER represents extent to which the foreground object occupies the road. Specifically, the road occupancy ratio increases with the increase in the sum of the MER areas.

Thus, road traffic occupancy σ is given by the following equation:

$$\sigma = \Big(\sum_{i=1}^{m} S'(d_i)\Big)/S, \tag{2}$$

where $S'(d_i)$ is the MER area of ith foreground d_i, S is the area of specified road in the image, and m indicates the number of foreground objects.

Because the road occupancy obtained by the CNN is based on detection results where it samples incoming frames once per second, we use a weighted smoothing method [24] for dynamic video processing to further improve the current occupancy value. Based on current σ_t and occupancy estimations from the previous three seconds, the improved occupancy can then be expressed as

$$\sigma_t = w_t\sigma_t + w_{t-1}\sigma_{t-1} + w_{t-2}\sigma_{t-2} + w_{t-3}\sigma_{t-3}, \tag{3}$$

where $w_t, w_{t-1}, w_{t-2}, w_{t-3}$ denote the weights of the current occupancy σ_t and weights of the occupancies from the previous three seconds respectively, and the weights satisfy $w_t + w_{t-1} + w_{t-2} + w_{t-3} = 1$. As the road status is a temporal feature, it means the current features are more closely related to the recently detected feature values, and the relationship with the past detected features are getting smaller and smaller as time goes by. We thus set these weights as an arithmetic progression to 0.49, 0.33, 0.17, 0.01 in our experiments.

B. Traffic Flow Detection

In the previous part, we discussed the detection of road occupancy. In practice, however, a single road occupancy characteristic cannot reflect road congestion status in all directions.

To be specific, the occupancy measure of a road on which a large van is traveling may be higher than the normal occupancy threshold while the road is under light traffic conditions.

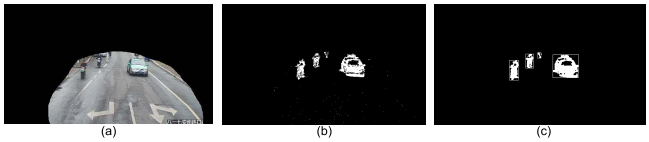

Fig. 3. An example of morphological processing for the foreground mask. (a) Current image. (b) The foreground mask before processing. (c) The foreground mask after processing.

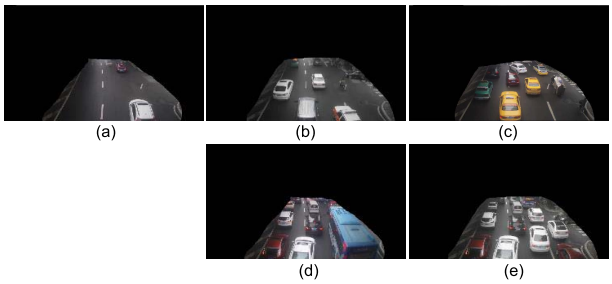

Fig. 4. An illustration of contrast character Con. The contrast increases gradually with greater vehicle density from 5.51 to 13.03. The higher contrast of the last panel where the road traffic is heavy is 2.37 times that of the first panel. (a) $Con = 5.51$. (b) Con = 7.01. (c) Con = 9.34. (d) Con = 11.81. (e) Con = 13.03.

Therefore, the visual feature of traffic flow can effectively compensate for the above parallel special case and diminish the false detection of road congestion caused by bias from the road occupancy measure for more accurate road congestion status.

Traffic flow refers to the number of vehicles passing through a specified section of road during a specified period of time. In this study, the road traffic flow and road occupancy obtained are both based on a CNN that samples incoming frames once per second. The final foreground object acquired from the CNN step can be quantitatively analyzed. That is, the occurrence likelihood of heavy traffic will increase in relation to a greater number of interesting foreground objects, and vice versa.

In our model, we get the correct objects of interest according to the CNN object detection by screening out false objects detected by the GMM. Then, each final foreground morphologically forms a connected area by post-processing so that the MER of each foreground object can be obtained as shown in Fig. 3. Next, we utilize bounding boxes to represent the MERs of connected foregrounds d_i. Every additional bounding box in the detected frame in our model means a new increment in the number of vehicles during the period of estimation for traffic flow.

Based on the discussion above, κ represents the traffic flow feature.

$$\kappa = |d_i \in D|, \tag{4}$$

where D and $|\cdot|$ denote the final foreground object set acquired during the CNN step and the number of set elements, respectively.

Here, we also use a weighted smoothing method [24] to further improve the current traffic flow. Based on the current κ_t and traffic flow estimations from the previous three seconds, the improved traffic flow can then be expressed as

$$\kappa_t = w_t\kappa_t + w_{t-1}\kappa_{t-1} + w_{t-2}\kappa_{t-2} + w_{t-3}\kappa_{t-3}, \tag{5}$$

where $w_t, w_{t-1}, w_{t-2}, w_{t-3}$ denote the weight of the current κ_t and the weights of the previous three-second traffic flow, respectively, which likewise can be set to 0.49, 0.33, 0.17, 0.01.

C. Traffic Density Detection

We have used two dimensions of the visual features to describe the traffic congestion. However, these two features alone are still not enough to comprehensively describe the complex condition of traffic congestion. Note that traffic flow and road occupancy rate both depict overall road character with global features in the foreground region, which reflect the relationship between the foreground object and the overall road. To investigate the relationship between vehicles and pedestrians, road traffic density features are considered to detect degree of congestion.

In our model, we take GLCM [25], [26] as a density clue to detect the traffic density on road since traffic density can be expressed with texture features based on GLCM via counting frequency of occurrence of image pixel pairs. Given an image $I \in \mathbb{R}^{M\times N}$, the GLCM is defined by the joint probability of density distribution of pixels $I(x_1, y_1)$ and $I(x_2, y_2)$ as follows:

$$M_{glcm}(i, j) = |I(x_1, y_1) = i, I(x_2, y_2) = j|, \tag{6}$$

$$(x_2, y_2) = d \cdot \theta + (x_1, y_1), \tag{7}$$

where $\{(x_1, y_1), (x_2, y_2) \in M \times N$, and N_g is the gray level of the input image, and the $M_{glcm}(i, j)$ and $|\cdot|$ represent the value of GLCM at position (i, j) and the number of elements in a particular set, respectively. d and θ denote the offsets and directions for (x_2, y_2), respectively.

After computing the GLCM, we do not directly use GLCM as a feature generally [27], [28] but extract the relevant statistical characteristics based on GLCM in order to speed up the calculation. We consider employing the expectation of the contrast feature Con of the GLCM in four directions to mirror the density of foreground objects. The final density feature Con can be calculated as follows:

$$Con' = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(i-j)^2 M_{glcm}(i, j), \tag{8}$$

$$Con = E\left(Con'\right), \tag{9}$$

where Con' denotes the contrast in each direction. The contrast character of GLCM is an indicator for contrast between foreground and background as well as an estimator of the object's density, as indicated in Fig. 4.

D. Traffic Velocity Detection

It is noted that the CNN-based occupancy detection and flow detection as well as the GLCM-based density detection all fall into the category of static visual feature detection. We further introduce traffic velocity detection for dynamic detection in

our method due to the consideration that our objects of interest are moving objects, and that the traffic may be flowing freely when the mentioned static visual features are large on special circumstances.

Our proposed method adopts the pyramid Lucas-Kanade optical flow method [17], [29], [30] to achieve detection of road traffic velocity because of its low computational cost and high robustness even under noise. With the combination of Lucas-Kanade optical flow and pyramid algorithm, we quickly get the moving speed of a foreground pixel by finding its corresponding spatial position in the time domain. Constant intensity and temporal continuity are both assumed in the Lucas-Kanade algorithm as shown below:

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t). \tag{10}$$

The Taylor expansion function of Eq. (10) is

$$I(x + \Delta x, y + \Delta y, t + \Delta t) \approx I(x, y, t) + I_x \Delta x + I_y \Delta y + I_t \Delta t, \tag{11}$$

where I_x, I_y, I_t is given by the derivative of I with respect to x, y, t. Then, we have

$$I_x \frac{\Delta x}{\Delta t} + I_y \frac{\Delta y}{\Delta t} = -I_t. \tag{12}$$

Further, the Lucas-Kanade method assumes the consistency of intensity change for neighboring pixels that are located in a sliding window of size $m \times m$ centered on the current pixel p. Let $\frac{\Delta x}{\Delta t}$ and $\frac{\Delta y}{\Delta t}$ denote V_x and V_y respectively. Therefore, we can rewrite Eq. (12) as

$$\begin{cases} I_{x_1} V_x + I_{y_1} V_y = -I_{t_1} \\ I_{x_2} V_x + I_{y_2} V_y = -I_{t_2} \\ \vdots \\ I_{x_n} V_x + I_{y_n} V_y = -I_{t_n}, \end{cases} \tag{13}$$

and $n = m^2$. Hence, we have $Av = b$, where

$$A = \begin{bmatrix} I_{x_1} & I_{y_1} \\ I_{x_2} & I_{y_2} \\ \vdots & \\ I_{x_n} & I_{x_n} \end{bmatrix}, \quad v = \begin{bmatrix} V_x \\ V_y \end{bmatrix}, \quad b = \begin{bmatrix} -I_{t_1} \\ -I_{t_2} \\ \vdots \\ -I_{t_n} \end{bmatrix}. \tag{14}$$

To make this equation well posed, the least squares method is used when $A^T A$ is non-singular. The traffic velocity v can be acquired as follows:

$$v = \left(A^T A\right)^{-1} A^T b. \tag{15}$$

In order to reduce the computational complexity, we only track corner pixels instead of all pixels in the image [31]–[33]. Finally, we take the expectation v^* of optical flow v_i for ith tracked corner pixel as traffic velocity as follows:

$$v^* = E(v) = \frac{1}{k} \sum_{i=1}^{k} v_i, \tag{16}$$

where k is the number of corner pixels.

IV. Multidimensional Traffic Congestion Detection Based on Visual Features and a CNN

A. Multidimensional Detection Model

Based on the above discussion, we finally obtain four visual features that can be used to describe road congestion status, including road traffic occupancy σ, traffic flow κ, traffic density Con, and traffic velocity v^*. We already know that road traffic is likely to congest when road occupancy σ and traffic flow κ as well as traffic density Con are large, but traffic velocity v^* is small.

Next, we investigate how to accurately detect road congestion status with visual features in different dimensions. The procedure of the proposed approach is shown in Fig. 1. In order to detect the congestion status of the road comprehensively, the congestion coefficient, which combines the discussed visual features, is given by following formula:

$$\varsigma = (w_{con} Con + w_\sigma \sigma + w_\kappa \kappa + \varepsilon)/\log v^*, \tag{17}$$

where w_{con}, w_σ and w_κ represent the weights of the corresponding feature, and ε denotes the error adjustment factor for the congestion coefficient ς.

However, due to the camera displacements, some error may occur in calculating the optical flow v^*. We use the standard HOF algorithm which is robust to enhance the accuracy and reliability of road congestion detection. When bin=30, the best results have been obtained. And the results of the experiment can be found that the value of bins is relatively stable to the HOF algorithm, and does not rapidly decay. The HOF can be constructed via accumulating optical flow within bins by the following formula:

$$v_i = (v_{i,x}, v_{i,y}), \tag{18}$$

$$\theta = \tan^{-1}(v_{i,y}/v_{i,x}), \tag{19}$$

$$\pi \frac{b-1}{bins} \le \theta \le \pi \frac{b}{bins}, \quad 1 \le b \le bins, \tag{20}$$

where b and θ are the bins' index in the HOF for current optical flows, and direction of the current optical flow, respectively. $bins$ represents the number of bins. Considering the length limit of the paper, we did not discuss HOF in detail, so in order to facilitate readers to understand the detailed steps, we gave a reference to the HOF method [34]. Then, We propose a method that measures the information entropy based on HOF. Hence, the histogram information entropy is expressed as:

$$Ent = \sum_{b=1}^{bins} -H(b) \log H(b), \tag{21}$$

where $H(b)$ denotes the value in bth bin. We find that the entropy of the HOF is small when there are shared directions for the majority of optical flows as desired, while the entropy is large when there are some other unexpected disturbances in optical flow yielded by camera displacements.

To integrate the entropy of the HOF to eliminate the error in optical flow, we have the final road congestion coefficient

ALGORITHM 2 Multidimensional Congestion Detection

Input: A captured sequence of road surveillance video D, $\varsigma_{\max}$, $\varsigma_{\min}$.

Output: The traffic status S.

```
for curframe = 0:totalframe do
  Label the segment of road to detect;
  Obtain F via GMM;
Post-process
  F in morphology;
  Detecte F using CNN by Algorithm 1;
  if curfram%fps==0 then
    Obtain traffic occupancy σ via Eq. (2);
    Obtain traffic flow κ via Eq. (4);
    Obtain traffic density Con via Eq. (6)-(9);
    Obtain traffic velocity v* via Eq. (10)-(16);
    Obtain congestion coefficient ς via Eq. (22);
    Detect the traffic status S via Eq. (23);
  end if
end for
```

as follows:

$$\varsigma = (w_{con}Con + w_{\sigma}\sigma + w_{\kappa}\kappa + \varepsilon)/(\log v^{*} - w_{ent}e^{Ent}), \tag{22}$$

where w_{ent} is the weight for entropy of the HOF. In doing so, the road congestion status is efficiently detected with the road congestion coefficient ς as follows:

$$S = \begin{cases} \text{“congestion”} & \text{if } \varsigma > \varsigma_{\max} \\ \text{“slowly”} & \text{if } \varsigma_{\max} \geq \varsigma \geq \varsigma_{\min} \\ \text{“free”} & \text{if } \varsigma < \varsigma_{\min}, \end{cases} \tag{23}$$

where $\varsigma_{\max}$ and $\varsigma_{\min}$ denote the upper and lower bounds for congestion detection, which can be learned from a number of training samples.

The following Algorithm 2 summarizes the multidimensional road congestion detection method proposed in this paper with the integration of visual features and CNN.

B. Parameter Selection

CNN is used to accurately detect the final foreground from the candidate foregrounds. During the training phase, the accuracy increases when the number of iterations is increased, but the problem of overfitting will appear when the number of iterations continues to increase. On one hand, the loss will be very high if the learning rate is high, and the model will oscillate repeatedly, eventually causing the model to diverge. On the other hand, the loss does fall if the learning rate is low, but the decline is too slow. Although the momentum factor cannot improve the accuracy, it can speed up the convergence rate. Weight decay is neither to improve the accuracy of convergence nor to increase the speed of convergence, and its ultimate goal is to prevent over fitting. Here, we take the ReLu (Rectified Linear Units) function as activation function and set the maximal number of iterations and base learning rate to 80000 and 0.001, respectively. Simultaneously, the learning weight decay is set to 0.0005 and momentum is 0.9 in our model. Experimental results show that the above parameters can achieve good results in foreground objects detection. The above deep learning parameters are used in the training phase, the accuracy will not change drastically if the parameter values changing.

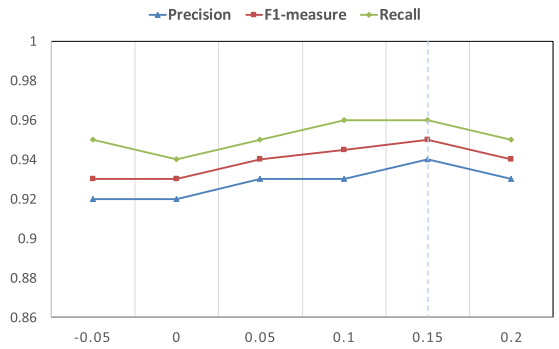

Fig. 5. Precision, F_1 and Recall as functions of w_{ent} with $w_{con} = 0.1$, $w_{\sigma} = 0.9$, $w_{\kappa} = 0.067$ and $\varepsilon = -0.5$. The F_1 is best when $w_{ent} = 0.15$.

Actually, the parameters w_{ent}, w_{con}, w_{σ}, w_{κ} of Eq. (22) restrict the effect of corresponding features, and the error adjustment factor ε is attributed to the shrinkage effect of the detection error. Therefore, the correct weights are important to determining traffic congestion and ensuring the accuracy of congestion detection. In our model, we need to properly normalize the features of different dimensions and thus infer the correct weights to evaluate the visual features of each dimension. Specifically, the ideal range for road traffic occupancy σ is in the [0, 1] interval, thus we set the weight $w_{\sigma} = 0.9$ in the experiment. Road traffic occupancy σ is a continuous value. It is found by experiment that the value of σ does not appear sensitive. The expectation of the GLCM contrast feature in four directions is used to mirror the density of foreground objects. Density feature Con is a continuous value. In the most congested case, the maximum of Con for density detection is 15, i.e., $Con_{\max} = 15$, and the minimum of Con for density detection is 5 when traffic is moving freely as shown in Fig. 4, i.e., $Con_{\min} = 5$. As the traffic density is directly connected to the intuitive performance of road congestion via the previous analysis, we let $(Con - Con_{\min})/10 = 0.1Con - 0.1Con_{\min}$ to get its weight $w_{con} = 0.1$ and $\varepsilon = -0.1Con_{\min} = -0.5$.The value of σ is stable to the experimental results, and the calculation process of density detection does not appear sensitive. In addition, the maximum value of traffic flow κ is $\kappa_{\max} = 15$. We hence set its weight to $w_k = \frac{1}{15} \approx 0.067$ for normalization. By fixing w_{con}, w_{σ}, w_{κ} and also fixing $\varsigma_{\max}$and $\varsigma_{\min}$ for congestion detection through experiments, we find after our analysis and reasoning that the congestion detection result is best where the weight of entropy $w_{ent} = 0.15$, as shown in Fig. 5.

Fig. 5 shows the effects of w_{ent} on each evaluation criteria of the experimental results. We can see that the overall evaluation rises steadily, and then declines gradually, and the F_1 is best when $w_{ent} = 0.15$. The Precision, F_1 and Recall results of the experiment change continuously with the value of w_{ent}. There is no sensitive point in the congestion detection algorithm.

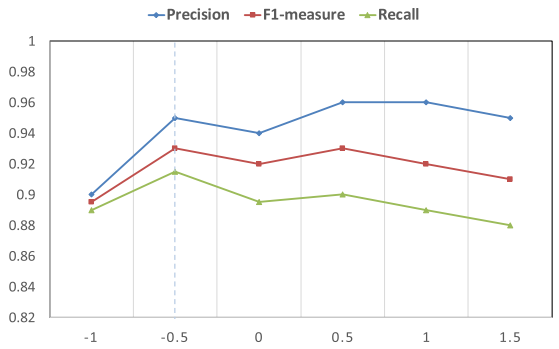

Fig. 6. Precision, F_1 and Recall as functions of ε with $w_{ent} = 0.15$, $w_{con} = 0.1$, $w_\sigma = 0.9$, $w_\kappa = 0.067$. The F_1 is best when $\varepsilon = -0.5$.

TABLE II
INFORMATION OF 25 TESTING REAL-WORLD TRAFFIC SURVEILLANCE VIDEOS CAPTURED IN DIFFERENT SCENARIOS

The information of 25 testing traffic surveillances			
Resolution	FPS	Average number of frames	Average length
1280 × 720	25 fps	≈1343 frames	≈53.7 second

The effects of ε on each evaluation criteria in our experiment is also presented in Fig. 6. As indicated in Fig. 6, we see that each evaluation increases sharply when ε changes from -1 to -0.5, and then presents a declining trend with an increasing difference between measures after ε is larger than -0.5.

Experimental results demonstrate that the detection result of the experiment is best, and the performance trend of the congestion coefficient ς can best express the change in traffic status when $w_{ent} = 0.15$, $w_{con} = 0.1$,$w_\sigma = 0.9$, $w_\kappa = 0.067$ and $\varepsilon = -0.5$.

V. EXPERIMENTAL RESULTS

In this study, the experiment is conducted on a Visual Studio 2013 platform with the C++ programming language. Due to the detection of traffic-congestion based on surveillance video is a new topic for study, there is thus not benchmark dataset for video-based congestion detection. Besides, it is very difficult to get the dataset of real-world traffic surveillance for congestion detection. Fortunately, our experiment video data is supported by telecommunication entity in China, which is permitted to set surveillance cameras in public traffic road. And our approach is currently focused on urban expressways and urban roads.

We test our method by labeling 3×33 sections of road according to the different traffic directions in 25 real-world traffic surveillance videos with 31795 frames in total, which are captured in 25 different scenarios. Fig. 7 displays the 25 scenarios of testing surveillance videos in our experiments to evaluate the performance of the proposed method. Table II presents detailed information about these testing videos, including resolution, frames per second (FPS), and the average number of frames and video length for each surveillance.

In this section, we present the quantitative and qualitative measurement conducted using our captured surveillance videos. By applying the proposed model, traffic congestion can be accurately described and detected from videos captured in different scenarios, as demonstrated by our experimental results. For these reasons, the performance of our proposed method has greatly surpassed that of its counterpart in our experiments.

Fig. 7. The 25 scenarios of testing surveillance videos, which are captured in different locations provided by a telecommunications entity in China.

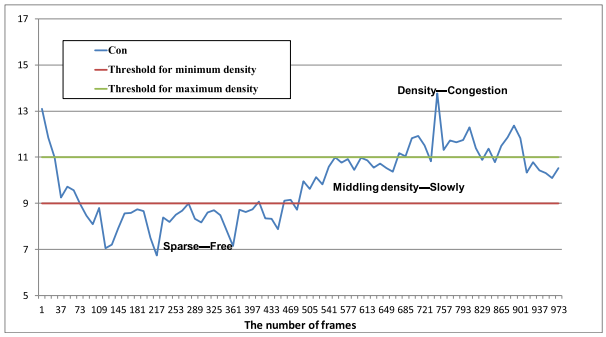

Fig. 8. Contrast features Con from an example road surveillance video with 960 frames in total.

A. Qualitative Assessment

1) Density Detection: We sample images for feature extraction once per second, and set d to 1 for density detection. We then calculate the feature of density Con after acquiring GLCM for each direction θ. There is a positive correlation between the contrast and the road density as demonstrated in Fig. 8.

Fig. 8 shows an example of video sequences where the contrast ranges from 6.72 to 13.76. Note that foreground objects are sparse since the image contrast is below the threshold for minimum density in the figure where traffic is moving freely. Contrast then varies steadily within a certain range in the upward process when foreground objects are less sparse (i.e. middling density) in this test video from the 505th frames. After that, road traffic gradually become congested,

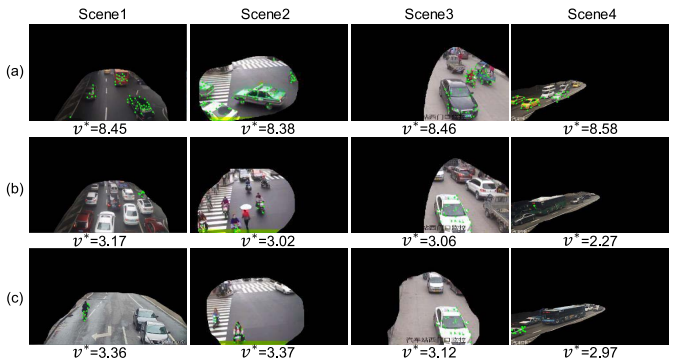

Fig. 9. Results of the optical flow for four subsequences of different scenes. (a) Free-flowing traffic with large optical flow. (b) Traffic congestion with small optical flow. (c) Free-flowing traffic with small optical flow.

and achieves peak congestion in the 745th frame. Accordingly, the contrast feature is in accord with the density variation of foreground objects on the road. That means that the contrast feature extracted by the GLCM is a favorable representation of traffic density detection.

However, it is not enough to learn the congestion status of road traffic using only the static feature of density. In some special cases, the traffic is still under free-flowing conditions even if the vehicle density is large while the average speed of traffic is fast.

2) Velocity Detection: The Lucas-Kanade optical flow is calculated iteratively using three-layer pyramid images. The size of the search window for each pyramid image is defined as 21×21 under resolution 1280×720. First, a set of traced corner pixels are detected [31], [32], and the Lucas-Kanade optical flow of the tracking corner is calculated using the information from the two adjacent frames. Here, we set the maximum number of corners to 500.

In our experiment, optical flow involving both length and direction is represented by green lines, and the end points are indicated by dots as shown in Fig. 9. Fig. 9 shows the results of four scenarios using the optical flow with pyramid method. When traffic is moving freely, the optical flow that corresponds to traffic velocity is large as show in Fig. 9 (a). Likewise, it is obvious that vehicles on the road are crowded, and the road is under congestion conditions in Fig. 9 (b) where the optical flow of vehicles is small. As a consequence, we can see that the optical flow feature is effective in traffic velocity detection.

3) CNN-Based Accurate Foreground Object Detection: However, in Fig. 9 (c) where the vehicles are stopped at a red light, the optical flow in Fig. 9 (c) is no larger than it is in Fig. 9 (b) while the road is clear and the vehicles are sparse. Consequently, traffic velocity is a necessary condition to describe traffic status, that is, from the velocity detection alone we still cannot fully describe and detect congestion status. In this part, we perform numerical experiments with regard to road occupancy and traffic flow, which are both calculated based on the CNN. Experiments on real video sequences demonstrate that foreground objects can be more accurately detected by integrating CNN and multidimensional visual features.

Consider the existence of shadows cast by the foreground candidates. Instead of radically removing these shadows, we simply leave them out of the calculation process for road occupancy and traffic flow to accurately detect the interesting foreground of a captured road surveillance video.

As a result, a CNN with an 11-layer core is constructed by virtue of the universality and efficiency of its feature maps. The feature map captures the main information of the features in the original image from different layers after convolution and sampling of the original image, including foreground, background, shape, textures, and other information. By doing so, the tedious process of feature selection is avoided, and we can finally classify the detected foreground objects accurately by using the CNN.

The classification with trained CNN is divided into three categories (i.e., vehicles, pedestrians or scooters, and others) in the experiment.

Fig. 10 shows the CNN-based foreground objects detection results of 15 scenarios. As show in Fig. 10, each scenario has two results, with one obtained directly by GMM without CNN, and the other detected GMM with CNN. In the detection results using GMM without CNN in Fig. 10, the green rectangle indicates the detected foreground. In the detection results by GMM with CNN, the green rectangle indicates the category of vehicles; while the blue rectangle indicates pedestrians or scooters, and a red rectangle represents a non-focused foreground such as shadows. It can be seen that a more accurate foreground detection can be achieved with a constructed and trained CNN. Moreover, the GMM-based CNN objects detection can facilitate the efficient detection for identifying the objects of small size with motion information based GMM, thus resulting in significant improvement in road occupancy σ and the traffic flow κ.

B. Quantitative Assessment

A quantitative assessment was conducted by evaluating the performance of each method when employed on our surveillance videos. We take K-fold cross-validation to test the robustness of the experiment. In our experience, K=10. Specifically, we divide the data set into 10 copies and take 9 of them as training data and 1 copy as test data in turn and we will get an accuracy rate each time. Lastly, we take the average accuracy of the 10 results as an estimate of the algorithm's accuracy. Additionally, three evaluation criteria are utilized to evaluate the performance of each compared approach in our experiments: Precision, Recall and F_1-measure. These criteria can be derived by assessing the detected road status of each compared algorithm as follows:

$$\text{Precision} = \frac{1}{n}\sum_{k=1}^{n}\frac{\text{TP}_k}{\text{TP}_k + \text{FP}_k}, \tag{24}$$

$$\text{Recall} = \frac{1}{n}\sum_{k=1}^{n}\frac{\text{TP}_k}{\text{TP}_k + \text{FN}_k}, \tag{25}$$

$$\text{F}_1\text{-measure} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}, \tag{26}$$

where TP_k is the number of correctly classified frames in kth status (i.e., the kth status can be congestion, slow, or free); FP_k

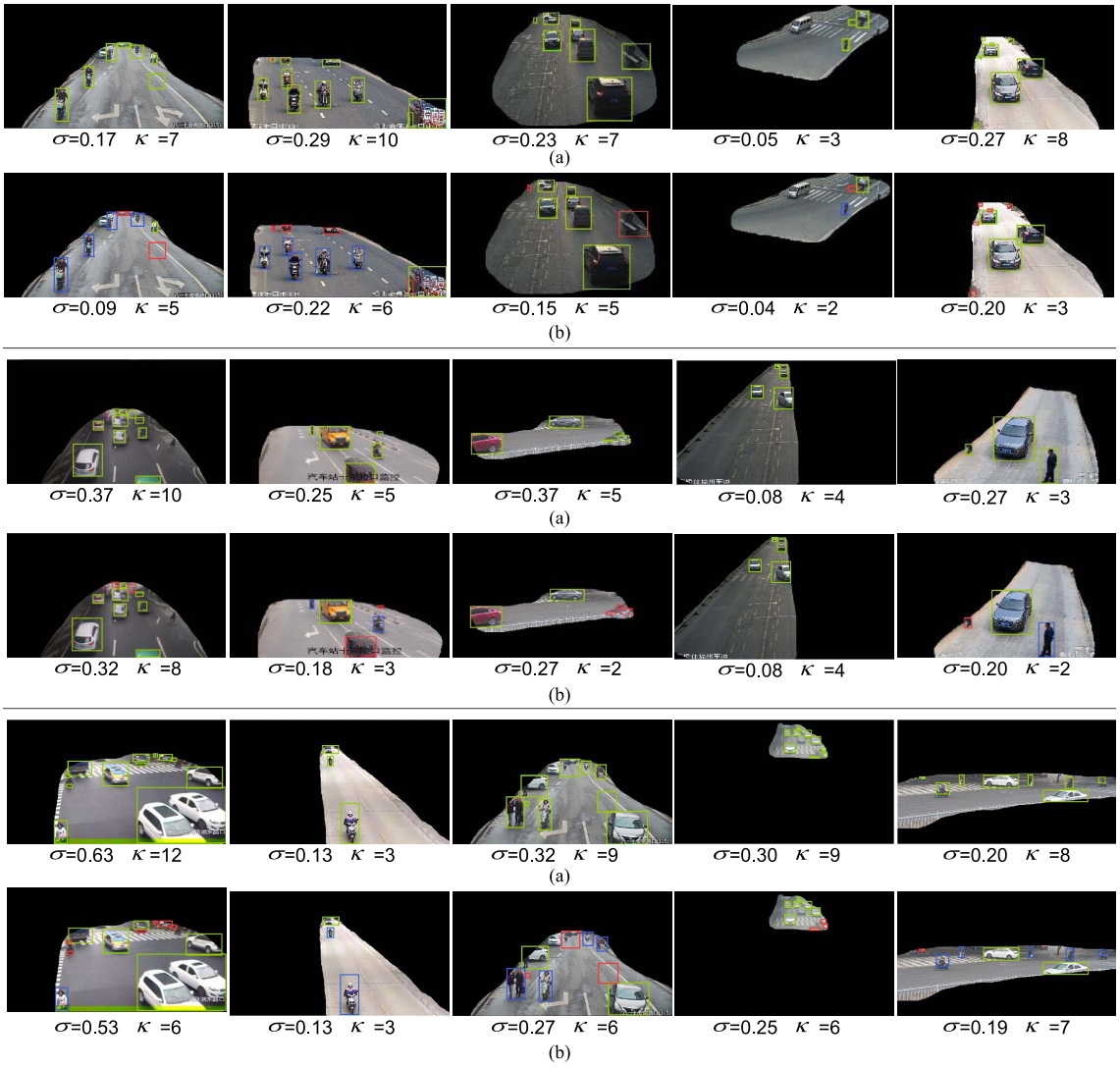

Fig. 10. Traffic occupancy σ and traffic flow κ of ten subsequences of captured surveillance video in 15 scenes which were obtained by the CNN for more accurate foreground detection. (a) The detection results by using GMM alone. (b) The detection results by using GMM-based CNN.

is the number of frames that are misclassified to kth status; TN_k is the number of correctly classified frames that were not in kth status; and FN_k is the number of frames in kth status but misclassified as not in kth status. Here, $n = 3$ denotes the three categories.

From the perspective of computer vision, we have calculated a continuous congestion factor ς by integrating multi-dimensional traffic factors (traffic density, traffic velocity, road occupancy and traffic flow), and measure traffic congestion from multiple dimensions. Therefore we define three states (congestion, slowly, free) to indicate traffic congestion. In our model, the upper and lower bounds for congestion detection $\varsigma_{\max}$ and $\varsigma_{\min}$ can be set to 1.5 and 0.5 respectively. Fig. 11 shows that ς changes with the variation in road status, and is intimately connected with road status.

In order to verify the effect of different dimensions of visual features to detect traffic status, we compare our method to five versions of our method with scaled-down features. The experimental results are displayed in Table III. As indicated

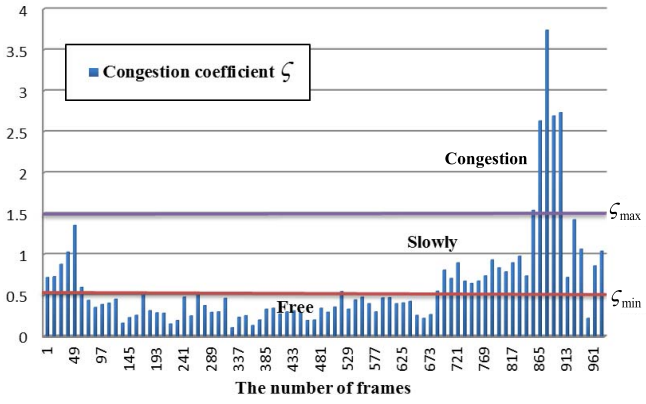

Fig. 11. An illustration of the congestion coefficient ς over a captured video sequence which contains three road statuses (i.e. congestion, slow traffic, and free-flowing traffic).

in Table III, Precision increased by 5 percent compared with the method without CNN (i.e., the 5th method in Table III). Furthermore, Precision, Recall, and the F_1-measure of the

TABLE III

COMPARATIVE RESULTS OF SIX VERSIONS OF OUR ALGORITHM BY DOWNSCALING THE VISUAL FEATURES WITH AND WITHOUT CNN (BEST PERFORMANCE IN BOLD)

	Method	Precision	Recall	F_1
1	GMM+(Con)	0.90	0.89	0.89
2	GMM+CNN+(κ,$v*$)	0.93	0.90	0.91
3	GMM+CNN+(Con,σ,$v*$)	0.94	0.92	0.93
4	GMM+CNN+ (Con,σ,κ)	0.95	0.93	0.94
5	GMM+(Con,σ,κ,$v*$)	0.91	0.89	0.90
6	Proposed: GMM+CNN+(Con,σ,κ,$v*$)	**0.96**	**0.94**	**0.95**

TABLE IV

DETECTION RESULTS FOR EACH COMPARED METHOD (BEST PERFORMANCE IN BOLD)

	Method	Precision	Recall	F_1
1	Marfia et al. [6]	0.92	0.91	0.91
2	Pan et al. [7]	0.91	0.90	0.90
3	Cao et al. [8]	0.93	0.90	0.91
4	Proposed method	**0.96**	**0.94**	**0.95**

proposed method are higher by 6%, 5%, and 6% respectively in comparison with the corresponding minimum result for each metric in the other methods.

Moreover, the feasibility of the proposed approach will be compared with three other state-of-the-art approaches based on their simulation results, including Marfia and Roccetti [6], Pan *et al.* [7], and Cao *et al.* [8]. Here, we mainly conduct experiments by quantitative measurements on our captured video data. Table IV reports the average accuracy attained by each criterion. It can be seen the proposed method can outperform other recent road congestion-status detection methods [6]–[8] by having the highest Precision (0.96), Recall (0.94), and F_1-measure (0.95), which is higher by 5%, 4%, 5%, respectively with respect to the lowest value in each metric.

This is due to both of [6] and [8] relying heavily on the veracity and quantity of relevant traffic information collected via hardware or manual operations, such as traversal times, entry and exit times, length of road, average size of vehicle, etc. In addition, [7] portrays the road-congestion status only by the linear relationship between the traffic velocity and the traffic density. In our method, a multidimensional road-status detection model with CNN is devised as indicated in Fig. 12. As a consequence, the accuracy of the proposed method is improved significantly by 5% in comparison with other recent methods [6]–[8] so that the status of roads can be detected more accurately.

Experiments demonstrate that the proposed method can effectively detect road congestion status, and validate the feasibility and effectiveness of our method by integrating static and dynamic features based on a CNN from a multidimensional feature space, including road occupancy, traffic flow, traffic density and traffic velocity.

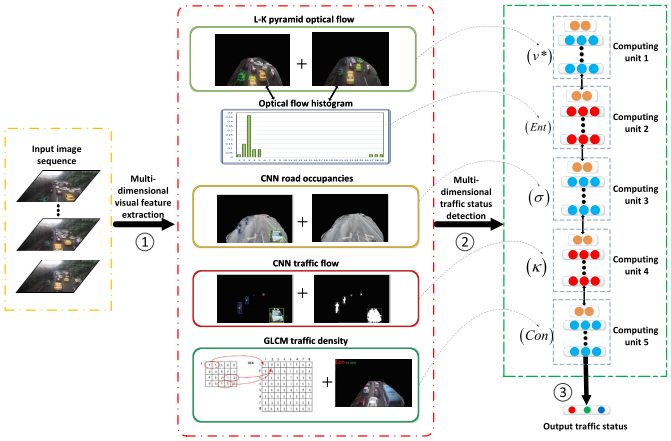

Fig. 12. Overview of the proposed approach. The proposed approach first extracts each visual feature, including traffic velocity $v*$, road traffic occupancy σ, traffic flow κ, and traffic density Con. Then, accurate traffic status detection can be achieved by integrating these visual features with the information entropy Ent of the HOF.

In addition, the average computational time in frames per second incurred by proposed method is 6.87 frames under resolutions 720p on a desktop computer with Pentium G3240-3.10GHz and 12GBRAM executing C++ codes. Considering the general performance of our experimental computer, when we replace CPU with I5 or I7, then the frame rate of our congestion algorithm can be even higher. Even under the current experimental conditions, for the normal 25FPS traffic video, the 6.87 frame of our algorithm can also achieve real-time detection of traffic congestion. Therefore, in the actual scenarios, our congestion algorithm has good scalability and feasibility.

VI. CONCLUSIONS

In this paper, a model for the effective detection of traffic congestion status on roads is presented. By carefully considering the factors influencing the status of congestion in a multidimensional feature space with a CNN, the proposed model is able to evade uninteresting foreground objects by constructing a CNN classifier, while at the same time extracting required visual features according to fine foreground pixels in a multidimensional feature space. Hence, the proposed method can yield better detection results with higher accuracy. The experimental results via quantitative and qualitative measurements show that the proposed method significantly outperforms other state-of-the-art methods with respect to Precision, Recall and F_1 metrics on our surveillance videos.

In the future, on one hand, we plan to further improve CNN for directly end to end detect the different traffic objects from road, including: cars, pedestrians and scooters. On the other hand, we plan to further study on how to employ the more effective motion information to CNN for dynamic video to get a practical application.

ACKNOWLEDGMENT

The authors gratefully acknowledge the helpful comments and suggestions of the reviewers.

REFERENCES

[1] R. Sun and Y. Chen, "Analysis of urban road congestion in China based on supply and demand perspective," in *Proc. Int. Conf. Remote Sens., Environ. Transp. Eng.*, Jun. 2011, pp. 69–71.

[2] X. Wang, "Intelligent multi-camera video surveillance: A review," *Pattern Recognit. Lett.*, vol. 34, no. 1, pp. 3–19, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S016786551200219X

[3] B. Tian *et al.*, "Hierarchical and networked vehicle surveillance in ITS: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 1, pp. 25–48, Jan. 2017.

[4] S. Sivaraman and M. M. Trivedi, "Vehicle detection by independent parts for urban driver assistance," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1597–1608, Dec. 2013.

[5] S. H. Khatoonabadi and I. V. Bajic, "Video object tracking in the compressed domain using Spatio-temporal Markov random fields," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 300–313, Jan. 2013.

[6] G. Marfia and M. Roccetti, "Vehicular congestion detection and short-term forecasting: A new model with results," *IEEE Trans. Veh. Technol.*, vol. 60, no. 7, pp. 2936–2948, Sep. 2011.

[7] J. Pan, I. S. Popa, K. Zeitouni, and C. Borcea, "Proactive vehicular traffic rerouting for lower travel time," *IEEE Trans. Veh. Technol.*, vol. 62, no. 8, pp. 3551–3568, Oct. 2013.

[8] Z. Cao, S. Jiang, J. Zhang, and H. Guo, "A unified framework for vehicle rerouting and traffic light control to reduce traffic congestion," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 7, pp. 1958–1973, Jul. 2017.

[9] F. Terroso-Saenz, M. Valdes-Vela, C. Sotomayor-Martinez, R. Toledo-Moreo, and A. F. Gomez-Skarmeta, "A cooperative approach to traffic congestion detection with complex event processing and VANET," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 914–929, Jun. 2012.

[10] J. W. Ju, J. Y. Liu, W. Yang, and N. Liu, "Modeling and simulation of vehicle traffic detection system based on ground sense coil," *Adv. Mater. Res.*, vols. 989–994, pp. 2511–2514, Jul. 2014.

[11] C.-H. Lo, W.-C. Peng, C.-W. Chen, T.-Y. Lin, and C.-S. Lin, "CarWeb: A traffic data collection platform," in *Proc. 9th Int. Conf. Mobile Data Manage. (MDM)*, Apr. 2008, pp. 221–222.

[12] S.-T. Jeng and L. Chu, "A high-definition traffic performance monitoring system with the inductive loop detector signature technology," in *Proc. 17th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2014, pp. 1820–1825.

[13] J.-W. Hsieh, S.-H. Yu, Y.-S. Chen, and W.-F. Hu, "Automatic traffic surveillance system for vehicle tracking and classification," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 2, pp. 175–187, Jun. 2006.

[14] M. Lei, D. Lefloch, P. Gouton, and K. Madani, "A video-based real-time vehicle counting system using adaptive background method," in *Proc. IEEE Int. Conf. Signal Image Technol. Internet Based Syst.*, Nov./Dec. 2008, pp. 523–528.

[15] J. Song, H. Song, and W. Wang, "An accurate vehicle counting approach based on block background modeling and updating," in *Proc. 7th Int. Congr. Image Signal Process.*, Oct. 2014, pp. 16–21.

[16] J. I. Engel, J. Martín, and R. Barco, "A low-complexity vision-based system for real-time traffic monitoring," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 5, pp. 1279–1288, May 2017.

[17] O. Barnich and M. Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1709–1724, Jun. 2011.

[18] W. Tao and K. Sun, "Asymmetrical Gauss mixture models for point sets matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1598–1605.

[19] N. Al-Najdawi, A. Abu-Roman, S. Tedmori, and M. Al-Najdawi, "An adaptive approach for real-time road traffic congestion detection using adaptive background extraction," *Int. Arab J. Inf. Technol.*, vol. 13, no. 3, pp. 327–335, 2016.

[20] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 1999, p. 252.

[21] H. Qin, J. Yan, X. Li, and X. Hu, "Joint training of cascaded CNN for face detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 3456–3465.

[22] A. Jourabloo and X. Liu, "Large-pose face alignment via CNN-based dense 3D model fitting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4188–4196.

[23] N. Seifnaraghi, S. G. Ebrahimi, and E. A. Ince, "Novel traffic lights signaling technique based on lane occupancy rates," in *Proc. 24th Int. Symp. Comput. Inf. Sci.*, Sep. 2009, pp. 592–596.

[24] N. T. Thomopoulos, *Applied Forecasting Methods*. Upper Saddle River, NJ, USA: Prentice-Hall, 1980.

[25] T. Torheim *et al.*, "Classification of dynamic contrast enhanced MR images of cervical cancers using texture analysis and support vector machines," *IEEE Trans. Med. Imag.*, vol. 33, no. 8, pp. 1648–1656, Aug. 2014.

[26] M.-C. Yang *et al.*, "Robust texture analysis using multi-resolution gray-scale invariant features for breast sonographic tumor diagnosis," *IEEE Trans. Med. Imag.*, vol. 32, no. 12, pp. 2262–2273, Dec. 2013.

[27] J. He and X. Zhu, "Combining improved gray-level co-occurrence matrix with high density grid for myoelectric control robustness to electrode shift," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 9, pp. 1539–1548, Sep. 2017.

[28] W. Gomez, W. C. A. Pereira, and A. F. C. Infantosi, "Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast ultrasound," *IEEE Trans. Med. Imag.*, vol. 31, no. 10, pp. 1889–1899, Oct. 2012.

[29] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. 7th Int. Joint Conf. Artif. Intell. (IJCAI)*, vol. 2. San Francisco, CA, USA: Morgan Kaufmann, 1981, pp. 674–679. [Online]. Available: http://dl.acm.org/citation.cfm?id=1623264.1623280

[30] X. Li, Y. She, G. Yang, Y. Zhao, and D. Luo, "A traffic congestion detection method for surveillance videos based on macro optical flow velocity," in *Proc. Int. Conf. Chin. Transp. Professionals*, 2015, pp. 1569–1578.

[31] J. Shi and C. Tomasi, "Good features to track," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 1994, pp. 593–600.

[32] M. S. Sigdel, M. Sigdel, S. Dinç, I. Dinc, M. L. Pusey, and R. S. Aygün, "Focusall: Focal stacking of microscopic images using modified Harris corner response measure," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 13, no. 2, pp. 326–340, Mar./Apr. 2016.

[33] P.-L. Shui and W.-C. Zhang, "Corner detection and classification using anisotropic directional derivative representations," *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3204–3218, Aug. 2013.

[34] J. Perš, V. Sulić, M. Kristan, M. Perše, K. Polanec, and S. Kovačič, "Histograms of optical flow for efficient representation of body motion," *Pattern Recognit. Lett.*, vol. 31, no. 11, pp. 1369–1376, 2010.

Xiao Ke received the Ph.D. degree in artificial intelligence from Xiamen University, Xiamen, China, in 2011.

He is currently an Associate Professor with the College of Mathematics and Computer Science, Fuzhou University, China. His current research interests include computer vision, pattern recognition, intelligent transportation, and machine learning.

Lingfeng Shi received the M.Sc. degree from the Department of Computer Science and Engineering, Yuan Ze University, Taoyuan, Taiwan, in 2017, and the M.Eng. degree from the College of Mathematics and Computer Science, Fuzhou University, China, in 2018.

Her current research interests include computer vision, intelligent transportation, and machine learning. She received the Outstanding Master Thesis Award from the Taiwanese Association for Consumer Electronics in 2017.

Wenzhong Guo received the B.S. and M.S. degrees in computer science and the Ph.D. degree in communication and information system from Fuzhou University, Fuzhou, China, in 2000, 2003, and 2010, respectively.

He is currently a Full Professor with the College of Mathematics and Computer Science, Fuzhou University. He currently leads the Network Computing and Intelligent Information Processing Laboratory, which is a key lab of Fujian Province, China. His current research interests include intelligent information processing, sensor networks, network computing, and network performance evaluation. He is a member of the ACM and a Senior Member of the China Computer Federation.

Dewang Chen (SM'14) received the Ph.D. degree in control theory from the Chinese Academy of Sciences in 2003. He was a Visiting Scholar with the Department of Electrical Engineering and Computer Science, University of California at Berkeley, in 2009.

He is currently a Minjiang Chair Professor with the College of Mathematics and Computer Science, Fuzhou University. He has authored over 60 papers and two monographs. His current research interests include intelligent control, machine learning, soft computing, and intelligent transportation systems.