



Research article



A deep neural framework for real-time vehicular accident detection based on motion temporal templates

Samy Bakheet^{a,b,*}, Ayoub Al-Hamadi^{b,**}^a Faculty of Computers and Artificial Intelligence, Sohag University, P.O. Box 82524, Sohag, Egypt^b Institute for Information Technology and Communications (IIKT) Otto-von-Guericke-University Magdeburg, D-39106 Magdeburg, Germany

ARTICLE INFO

Keywords:

Vehicular accident prediction
 Temporal templates
 Moment invariants
 Fuzzy time slicing
 Cubic-spline DNN

ABSTRACT

Vehicular accident prediction and detection has recently garnered curiosity and large amounts of attention in machine learning applications and related areas, due to its peculiar and fascinating application potentials in the development of Intelligent Transportation Systems (ITS) that play a pivotal role in the success of emerging smart cities. In this paper, we present a new vision-based framework for real-time vehicular accident prediction and detection, based on motion temporal templates and fuzzy time-slicing. The presented framework proceeds in a stepwise fashion, starting with automatically detecting moving objects (i.e., on-road vehicles or roadside pedestrians), followed by dynamically keep tracking of the detected moving objects based on temporal templates, clustering and supervised learning. Then, an extensive set of local features is extracted from the temporal templates of moving objects. Finally, an effective deep neural network (DNN) model is trained on the extracted features to detect abnormal vehicle behavioral patterns and thus predict an accident just before it occurs. The experiments on real-world vehicular accident videos demonstrate that the framework can yield mostly promising results by achieving a hit rate of 98.5% with a false alarm rate of 4.2% that compare very favorably to those from existing approaches, while still being able to deliver delay guarantees for realtime traffic monitoring and surveillance applications.

1. Introduction

A recent report by the World Health Organization (WHO) has indicated that annually approximately 1.35 million people die and between 20 and 50 million people are injured or disabled by non-fatal injuries worldwide as a result of vehicular accidents [1]. On the other side, road traffic plays a dominant role in our daily lives and a large variety of human activities and services are increasingly dependent, either directly or indirectly, upon it [2]. Effective smart vehicular-traffic management regarded as a precursor of the development of the so-called intelligent transportation systems (ITS) is pivotal to public safety in high traffic locations especially where there are unfamiliar conditions. This is particularly important for those in charge of traffic control and monitoring. The ability to estimate and forecast fatal vehicular accidents and provide specific key details such as when, where and how the accident occurred is particularly pivotal not only to public-safety stakeholders (e.g., police), but also to road travelers and transportation administrators.

With most existing traffic management systems, traffic monitoring and surveillance are commonly performed by means of fixed surveillance cameras installed by traffic police departments on several junctions and along the roads. However, a common practice of such traffic monitoring is predominantly performed semi manually by traffic police staffs sitting in a control room while monitoring the installed cameras. This not only involves a considerable amount of effort and time supplied by human operator, it also lacks support for desired real-time responsiveness to sudden/unexpected traffic events.

On the other side, advanced traffic surveillance and intelligent transportation systems that heavily rely on computer vision and machine learning algorithms currently constitute a trendy research topic where a diversity of promising approaches and methodologies are being pursued to detect, localize, track, and even recognize moving vehicles in video streams captured by roadside cameras in unconstrained traffic scenes with little or no significant human interventions [3, 4, 5, 6]. Moreover, such smart systems have great potential to monitor and evaluate on-road vehicle driving behaviors and eventually generate a fast and

* Corresponding author at: Faculty of Computers and Artificial Intelligence, Sohag University, P.O. Box 82524, Sohag, Egypt.

** Corresponding author at: Institute for Information Technology and Communications, Otto-von-Guericke-University, D-39106 Magdeburg, Germany.

E-mail addresses: samy.bakheet@fci.sohag.edu.eg (S. Bakheet), ayoub.al-hamadi@ovgu.de (A. Al-Hamadi).<https://doi.org/10.1016/j.heliyon.2022.e11397>

Received 18 October 2021; Received in revised form 19 January 2022; Accepted 28 October 2022



Fig. 1. Automatic vehicular accident detection and notification.

precise semantic interpretation based on the results of the vehicle motion characteristics (see Fig. 1). This, in turn, is expected to contribute significantly to vehicular traffic collision and/or vehicular congestion reduction by facilitating daily traffic control and management and allowing for an instant response when sudden vehicle events occur.

The objective of this paper is twofold: firstly, to investigate the problem of prediction and detection of vehicular accidents based on vehicle motion analysis and secondly, to propose a new vision-based framework for predicting and detecting vehicular accidents in real-time, based on motion temporal templates and fuzzy time slicing. Within this framework, moving objects (i.e., on-road vehicles or roadside pedestrians) are first detected and dynamically tracked based on motion temporal templates, clustering and supervised learning. A cubic-spline neural model is then trained on an optimum set of localized features extracted from predefined motion templates to predict an accident just before it occurs and a forward warning will alert the drivers to provide them the opportunity to take appropriate evasive action in time. Furthermore, non-straight trajectories of individual pre-accident vehicles can be obtained from vehicle tracking. This information recorded and stored automatically by the system can be crucial sources of evidence in the assessment and investigation of the traffic accident, by providing inclusive guidance for investigators in determining accident causes and follow-up action. The rest of this paper is organized as follows. In Section 2 we outline relevant prior work. Temporal motion templates are thoroughly covered in Section 3. In Section 4, the proposed framework is presented in details and its components are discussed. Experimental results are presented and analyzed in Section 5. Finally, in Section 6, we draw some conclusions and include suggestions for further work.

2. Prior work

During the last three decades or so, the research on automated detection of traffic accidents and driving violations has attracted extensive interest of a vast number of researchers in different fields such as neuroscience, computer vision and robotics. As a result, tremendous research efforts have been dedicated to deploying automated accident detection systems onto urban roads [7, 8]. In the literature, there are numerous approaches proposed for automatic traffic accident detection using a variety of approaches such as Kalman filters, decision trees, or time series analysis and forecasting, with differing levels of success rate in their performances [9, 10].

A most recent work that is closely related to ours in this area is perhaps that was presented in [11], where a machine learning framework using multimodal sensors in vehicles is proposed for automated accident detection. In that innovative study, five state-of-the-art feature extraction approaches such as methods relying on feature engineering and learning using deep learning were evaluated on the second strategic highway research program naturalistic driving study (SHRP2 NDS) dataset to detect realistic driving crashes. By the study results, it was observed that CNN features with a support vector machine (SVM) model outperform all other approaches considered. Moreover, unsupervised

feature extraction was shown to be remarkably effective in achieving a notable performance score. In a similar study performed by Jahangiri et al. [12], a random forest (RF) model is used for classifying driver behaviors into two main classes: violation and compliance at junctions with signals. In their work, the authors make use of the intersection's distance, speed, acceleration, time, necessary deceleration parameter, and velocity-based handmade features recorded by radar, video cameras, and signal phase sniffers at intersections. At signalized intersections, SVM and RF architectures could successfully predict driving violations, with accuracies of roughly 0.98 and 0.94, respectively.

In [13], the authors adopt an SVM for classifying driver activities based on hand postures acquired by cameras equipped in the vehicles, while in [14], an automatic system is proposed for recognizing car colors and logos, based on color segmentation and labeling. Moreover, in [15], a long-term prediction approach is introduced based on a combined trajectory classification and particle filter framework, where trajectories are classified using a rotationally invariant version of the longest common subsequence as a similarity metric between trajectories, with an architecture for processing trajectories of arbitrary non-uniform length. In a similar vein, in [16], a framework is proposed for learning multi-lane trajectories based on vehicle mounted vision sensors, where typical trajectories in highway driving are modeled using Kalman filtering and hidden Markov modeling is used for classification. Moreover, Wiest et al. [17] present a similar approach for predicting the long-term trajectories of vehicles, based on variational Gaussian mixture modeling, where the trajectory of vehicle is obtained several seconds in advance.

A variety of methods have also been proposed in the literature for detecting traffic accidents incorporating different type of cues such as matrix approximation [18], Smoothed Particles Hydrodynamics (SPH) [19], Scale Invariant Feature Transform (SIFT) and optical flow [20], or adaptive traffic motion flow modeling [21]. Due to the recent revolution in object detection based on deep learning, a great number of methods have been introduced for automatic traffic accident detection and anticipation, e.g., using Convolutional Neural Networks (CNNs) [22, 23] and Recurrent Neural Networks (RNNs) [7, 24], along with various disparate traffic accident datasets [7, 23, 24, 25, 26] available involving surveillance videos or dashcam videos. For example, in [27], the authors proposed a feature fusion deep learning approach for urban traffic crash detection, aiming at striking some kind of a feasible balance between detection time and accuracy with a minimum of computing resources. Despite their robustness, most of the methods mentioned above suffer from operational inefficiency and a lack of real-time performance in online accident detection without using future frames and they still yield unsatisfactory results. Additionally, there is no specialized dataset for traffic analysis that contains video streams with top-down views similar to drone/unmanned aerial vehicle videos, or omnidirectional camera videos.

3. Temporal templates

Strictly speaking, temporal templates are essentially 2D representations of motion history, created as 2D images from 3D video streams. They are unarguably perceived to be the most effective at reducing a 3D spatiotemporal space to a 2D image representation [28]. Exquisitely conformed with a cumulation of successive image differences, the temporal templates can potentially capture motion information by encoding where and when motion happened in the image sequences. Despite the removal of one dimension, the temporal information is mostly preserved in the associated 2D image. The construction process of temporal templates requires either a stationary camera and a relatively static background, or the segregation of the motion of target from the background motion.

In their pioneer work, Bobick and Davis [28] have introduced two types of temporal templates, namely Motion-Energy Image (MEI) and Motion-History Image (MHI). The MEI template is created without

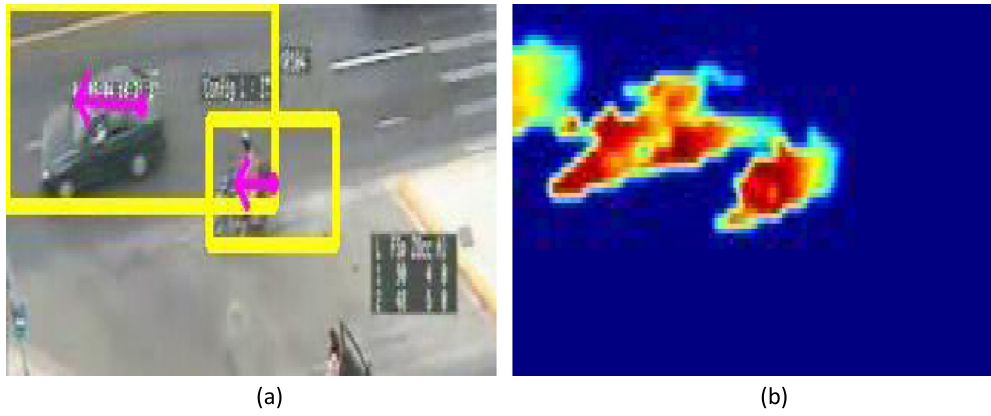


Fig. 2. A sample snapshot of a traffic video steam: (a) foreground motion detection and (b) created MHI temporal template.

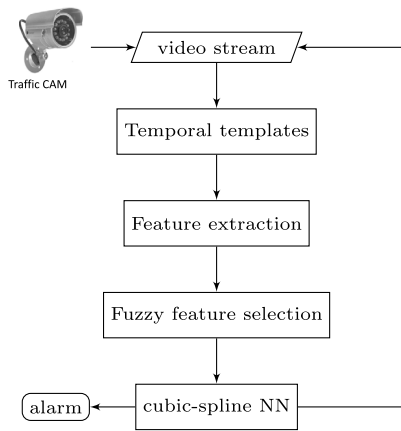


Fig. 3. A functional block diagram of the presented framework for traffic accident detection.

keeping the information about the time at which the motion takes place, while in the MHI template, the temporal information is collapsed into a single grayscale image, where the brightness of a particular pixel is directly proportional to the recent pixel motion. In Fig. 2, an example of MHI temporal templates for a traffic video steam is shown. A persistent challenge inherent to the original temporal template approach [28] is motion overwriting due to motion self-occlusion which remains a daunting task to solve for motion recognition. An avenue to best cope with this challenge is to build multilevel MHI (MMHI). It is specifically permitted to record the motion history at multiple time intervals, instead of recording it just once for the full video sequence and building a single MHI.

4. Proposed method

In this section, the proposed framework for real-time near-accident detection in traffic video streams is described. Our primary objective here is to develop a fast and reliable system that can automatically predict traffic accidents in advance. We take into account that the detection process of vehicle motion accommodates a variety of lighting conditions. Moreover, the proposed system is designed to be particularly powerful, quite, robust, and gains the best accuracy/speed trade-off. The block diagram representing the general outline of the suggested framework is depicted in Fig. 3. The framework can attain its intended goals by going through four subsequent steps described in the following subsections.

4.1. Temporal template formulation

Assume $I(x, y, t)$ denotes the luminance value which changes over time to produce the video sequence. In addition, suppose that $B(x, y, t)$ denotes the difference image describing the luminance variations and detected by a simple frame-differencing scheme: $B(x, y, t) = |I(x, y, t) - I(x, y, t + \delta t)|$, where x and y represent the 2D spatial coordinates in horizontal and vertical directions, respectively, and δ is the temporal distance. More formally, in an MHI denoted by H_τ where the parameter τ determines the temporal duration (in frames), the value of each point's intensity in this representation can be defined as a function of motion attributes at the corresponding image position in the video sequence. Consequently, $H_\tau(x, y, t)$ can be recurrently calculated with an update function $\psi(x, y, t)$ by a recurrence formula as follows:

$$H_\tau(x, y, t) = \begin{cases} \tau, & \text{if } \psi(x, y, t) = 1 \\ \max(0, H_\tau(x, y, t-1) - \epsilon), & \text{o.w} \end{cases} \quad (1)$$

where $\psi(x, y, t)$ and ϵ denote the moving object in the current image frame and the decay parameter, respectively. The MHI can be then obtained by thresholding $D(x, y, t)$ using Eq. (2) as follows,

$$\psi(x, y, t) = \begin{cases} 1, & \text{if } D(x, y, t) > \xi \\ 0, & \text{o.w} \end{cases} \quad (2)$$

where ξ denotes a certain threshold value. It is perhaps noteworthy here to mention that if the information about the beginning and end of the motion of interest is unavailable, it is perhaps desirable or even necessary to alter the period τ and then try to categorize the resulting MHIs [29]. Conversely, when the start and end of a motion pattern are determined and correspond to the duration of the recorded video stream, in this situation the parameter τ doesn't need to be varied. It is possible to appropriately normalize the temporal dynamics by the general distribution of the gray values within the MHI throughout the valid range (i.e., [0-255], with 8-bit grayscale levels). Consequently, the display duration of a motion unit can be consistently maintained. This allows identical motion patterns of different periods to be legitimately compared and contrasted with each other. Since the video sequences are composed of different number of frames, the number of history levels in MHIs might still differ from one sequence to another. To appropriately compare the video sequences, it is essential that the multi-level MHI (MMHI) approach allows all MHIs to be constructed with a fixed number ℓ of history levels. Video streams are then downsampled to $\ell + 1$ key image frames. Hence, taking into account the known parameter ℓ , the recursive MHI formula in Eq. (1) is extended to be as follows,

$$H(x, y, t) = \begin{cases} \alpha t, & \text{if } \psi(x, y, t) = 1 \\ H(x, y, t-1), & \text{o.w} \end{cases} \quad (3)$$

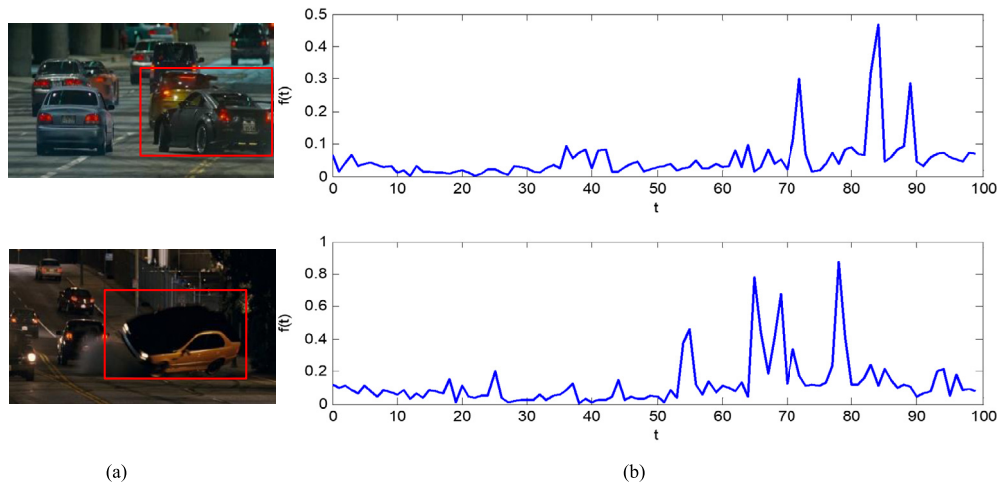


Fig. 4. Two sample vehicular accidents along with their MHI based features: (a) frame shots with accident regions boxed in red and (b) corresponding MHI based features.

where α denotes intensity difference between each two historical levels. It is noteworthy that Eq. (3) reveals that $H(x, y, t) = 0$ for $t = 0$. With an MMHI, the ultimate goal is to encode the motion occurrence at different time instants on the same pixel location such so that it can later be uniquely decoded. To accomplish this goal, we can use a simple bitwise coding scheme, where an additional term (2^{t-1}) is appended to the previous MMHI values, when motion takes place at an instant t and in a pixel position (x, y) :

$$M(x, y, t) = M(x, y, t-1) + 2^{t-1}\psi(x, y, t) \quad (4)$$

where $M(x, y, t = 0) = 0$. Thus far, the bitwise coding scheme as in Eq. (4) has the potential to deal effectively with the presence of multiple motions at the same location and temporally cluster them into individual motion patterns.

4.2. Feature extraction

This section comprises of the details of the feature extraction procedure based on temporal motion templates described previously. More specifically, we explain how a set of easily computable features (i.e., statistical and geometrical features) can be extracted from spatio-temporal motion templates of moving-vehicle sequences, and how these discriminative features can be used for traffic accident detection and prediction. The set of simple statistical features used in this work for representing vehicle actions to predict traffic accidents includes:

1. The mean direction of moving regions defined as the angle between the x-axis and the motion orientation by the clockwise, which is computed by a weighted-oriented histogram, in which a more recent motion has a greater weight than that assigned to the motion occurred early.
2. The shortest distance from the start-point to the endpoint of the motion trajectory at each time instant.
3. The filling rate of the bounding box surrounding the greatest region of interest (ROI) in MHI, which signifies the portion of the bounding box of motion taking place at a time instant.
4. The ratio of width to height of rect around the moving object, which can well describe the motion characteristics from shape aspects.

Since we are interested, in this work, to represent vehicle actions on more local level, then we move one step further and propose to extract an additional set of geometrical features that are most relevant for the current detection task.

To achieve this objective, we divide the bounding box for the region of detected motion spatially into equally-sized subregions. Thus, an additional set of affine geometric moment invariants that have good stability and distinguishability can be efficiently extracted from each partitioned motion subregion [30] using Eq. (5), as follows:

$$\begin{aligned}
 J_1 &= \frac{1}{\mu_{00}^4} [\mu_{20}\mu_{02} - \mu_{11}^2] \\
 J_2 &= \frac{1}{\mu_{00}^{10}} [\mu_{03}^2\mu_{30}^2 - 6\mu_{30}\mu_{21}\mu_{12}\mu_{03} + 4\mu_{30}\mu_{12}^3 + 4\mu_{03}\mu_{21}^3 \\
 &\quad - 3\mu_{21}^2\mu_{12}^2] \\
 J_3 &= \frac{1}{\mu_{00}^7} [\mu_{20}(\mu_{21}\mu_{03} - \mu_{12}^2) - \mu_{11}(\mu_{30}\mu_{03} - \mu_{21}\mu_{12}) \\
 &\quad + \mu_{02}(\mu_{03}\mu_{12} - \mu_{21}^2)] \\
 J_4 &= \frac{1}{\mu_{00}^{11}} [\mu_{20}^3\mu_{03}^2 - 6\mu_{20}^2\mu_{11}\mu_{12}\mu_{03} - 6\mu_{20}^2\mu_{02}\mu_{21}\mu_{03} \\
 &\quad + 9\mu_{20}^2\mu_{02}\mu_{12}^2 + 12\mu_{20}\mu_{11}^2\mu_{21}\mu_{03} + 6\mu_{20}\mu_{11}\mu_{02}\mu_{30}\mu_{03} \\
 &\quad + 18\mu_{20}\mu_{11}\mu_{02}\mu_{30}\mu_{12} - 8\mu_{11}^3\mu_{30}\mu_{03} \\
 &\quad - 6\mu_{20}\mu_{02}^2\mu_{30}\mu_{12} + 9\mu_{20}\mu_{02}^2\mu_{21}^2 + 12\mu_{11}^2\mu_{02}\mu_{30}\mu_{12} \\
 &\quad - 6\mu_{11}\mu_{02}^2\mu_{30}\mu_{12} + \mu_{02}^3\mu_{30}^3] \\
 J_5 &= \frac{1}{\mu_{00}^6} [\mu_{40}\mu_{04} - 4\mu_{31}\mu_{13} + 3\mu_{22}^2] \\
 J_6 &= \frac{1}{\mu_{00}^9} [\mu_{40}\mu_{04}\mu_{22} + 2\mu_{31}\mu_{13}\mu_{22} - \mu_{40}\mu_{13}^2 \\
 &\quad - \mu_{04}\mu_{13}^2 - \mu_{22}^3] \quad (5)
 \end{aligned}$$

where μ_{ij} denotes the normalized central moment of order $i + j$. In Fig. 4, sample snapshots from two different vehicular accident scenes along with their associated MHI based features are shown.

4.3. Fuzzy feature selection

This section is dedicated to the scheme utilized for feature selection based on adaptive decomposition an input vehicle sequence into a finite number of time slices in a fuzzy manner. Specifically, this feature selection scheme intends to select a feature subset from the extracted features by the elimination of redundant or irrelevant features. Hence, not only does the reduced set of relevant features yield a considerable speed-up for the prediction process of traffic accidents by pruning out irrelevant features, but it also has the potential to provide better (or at least the same) performance on prediction of traffic accidents as the initially extracted ones can achieve. Eventually, this allows the presented

approach to achieve the goal of feature dimension reduction, without a significant loss in final prediction accuracy. As described in the previous section, a feature vector can be simply generated at each time instance t by combining extracted features as follows,

$$\mathbf{f}_t = (f_{t,1}, f_{t,2}, \dots, f_{t,l})^T \quad (6)$$

where l is the number of features obtained at an instant time. As the extracted features in Eq. (6) are computed from a traffic image sequence (i.e., vehicle action snippets) at each time instance, each vehicle action can then be expressed by a time series of the features: $\{\mathbf{f}_t\}_{t=0}^{T-1}$ that give us a rigorous approach for detecting vehicle actions. To calculate the final feature vector for a vehicle action, the video snippet of each vehicle action is temporally divided into multiple time-slices defined by linguistic intervals [31]. Each of these intervals is described by a Gaussian fuzzy membership function that is expressed as follows:

$$\psi_j(t; \mu, \sigma, \xi) = e^{-\left|\frac{t-\mu}{\sigma}\right|^\xi} \quad (7)$$

where μ, σ and ξ in Eq. (7) are tunable parameters of the Gaussian function. Hence, we can now generate a distinctive feature vector for each time-slice, by computing a weighted average of features in all frames of the time-slice [32, 33]. More formally, we can create a feature vector for i th time-slice as follows:

$$\mathbf{F}_i = \frac{1}{\Delta t} \sum_{t \in slice_i} \psi_i(t) \mathbf{f}_t, \quad i = 1, 2, \dots, m \quad (8)$$

where $\psi_i(t)$, Δt and m are the Gaussian function representing the i -th time-slice, the duration of the time-slice in frames, and the total number of time-slices of the vehicle video snippet, respectively. Hence, the final feature vector for a given vehicle action can be directly created by concatenating all the feature vectors in Eq. (8) as follows: $F = \cup_{i=1}^m \mathbf{F}_i$, where \cup represents the concatenation of feature vectors.

4.4. Feature classification

This section gives the details of the feature classification module in the proposed vehicular-accident detection framework. Broadly speaking, there are numerous machine learning algorithms [34, 35] that can train a detector of vehicular accidents. Due to its superior intrinsic generalization capabilities and reputation as a robust and accurate paradigm, an adaptive DNN model is employed for the current classification task, which consists of several hidden layers of artificial neurons that are fully connected with neurons of the previous and following layers. In the digital implementations of neural networks, the activation function that intuitively simulates the behavior of biological neurons is basically used as a decision making body at the output of neurons. In DNN implementations, a wide variety of well-known activation functions including sigmoid, Gaussian, step, linear, piecewise, and tangent hyperbolic functions are widely used to implement the neuron behavior. To promote the flexibility and learning ability of the neural classification model adopted in this work, an adaptive activation function out of a pool of well-known functions is employed for the hidden neurons, namely a cubic-spline function. It should be kept in mind that the proper choice of activation function should consider some well-known aspects such as easy implementation, efficient computation, and the utilized function should be partially refinable that means it can be replaced with smaller copies of itself at regular intervals. These requirements lead us to a mathematical field concerned with interpolating polynomials through given data points, a part of numerical analysis. A cubic spline function is defined as a third degree piecewise polynomial, which yields a smooth curve that passes through all of the data points (i.e. knots). More formally, given a data set $(x_j)_{j=0}^n$ of $n + 1$ points, a cubic-spline function can then be expressed as follows,

$$S(x) = \cup_j s_j(x) = \cup_j \sum_{k=0}^3 a_{j,k} (x - x_j)^k \quad (9)$$

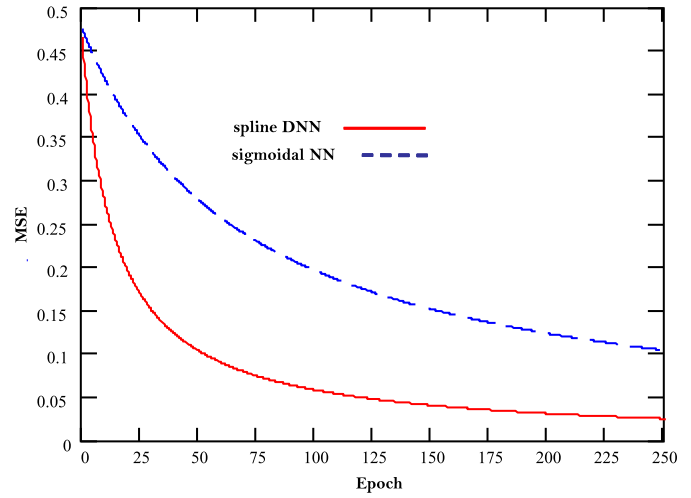


Fig. 5. Learning curve comparison between the cubic-spline DNN model and the standard sigmoidal neural model.

$$\forall x \in [x_j, x_{j+1}], j = 0, 1, \dots, n - 1$$

where $a_{j,k}$ and \cup are the coefficients of cubic-spline function and concatenation operator, respectively. To evaluate the above coefficients, a coefficient matrix of a linear system of four equations in four unknowns is constructed using general characteristics of spline functions such as interpolating property, twice continuous differentiable, etc). Therefore, by using Eq. (9) and splines' properties (for additional information, including calculation details, refer to [36]), the following matrix system of equations holds:

$$\begin{pmatrix} a_{j,0} \\ a_{j,1} \\ a_{j,2} \\ a_{j,3} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -\alpha & 0 & \alpha & 0 \\ 2\alpha & \alpha - 3 & 3 - 2\alpha & -\alpha \\ -\alpha & 2 - \alpha & \alpha - 2 & \alpha \end{pmatrix} X, \quad (10)$$

$$\text{where } X = \begin{pmatrix} x_{j-1} \\ x_j \\ x_{j+1} \\ x_{j+2} \end{pmatrix}, j = 1, 2, \dots, n - 1$$

where α represents the tension parameter in the cubic-spline function, which defines how sharply the spline curves bend at the control points. This parameter is commonly set to 0.5. Performing the substitution $\alpha = 0.5$ in Eq. (10) ultimately results in the following expression (Eq. (11)) for the cubic-spline function:

$$s_j(u) = \frac{1}{2} (1 \quad u \quad u^2 \quad u^3) AX, \quad (11)$$

$$\text{where } A = \begin{pmatrix} 0 & 2 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 2 & -5 & 4 & -1 \\ -1 & 3 & -3 & 1 \end{pmatrix}, 0 \leq u \leq 1$$

The averaged learning curves for the cubic-spline DNN model and standard sigmoidal neural model are shown in Fig. 5. A cursory glance on this figure clearly shows that the modified cubic-spline neural model not only has a much faster convergence than the standard sigmoidal neural model at the early stage of training, but also constantly excels during the rest of the training process.

5. Experimental results

In this section, the experiments performed to evaluate and validate the performance of the proposed framework are described and the obtained results that demonstrate the effectiveness and efficiency of the framework are presented and discussed. Due to the intense difficulty

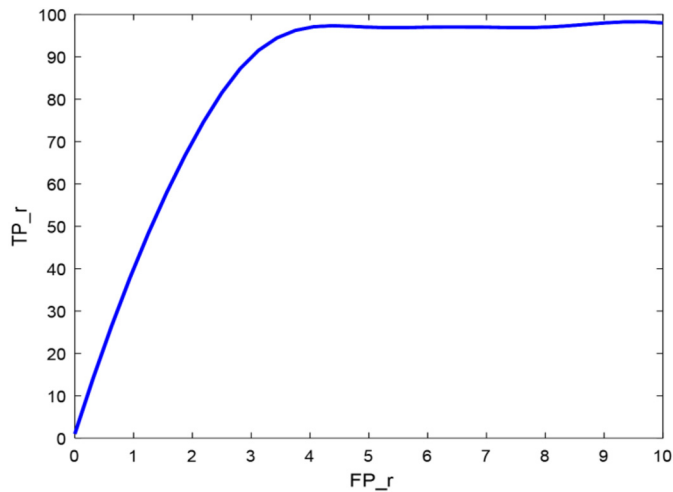


Fig. 6. ROC curve of the proposed system for vehicular accident prediction and detection.

(and inherent danger) involved in capturing or simulating abnormal traffic events (e.g., traffic accidents and violations) in natural driving scenes, it was only possible to perform a series of experiments with a relatively limited dataset of realistic accident scenarios. We have evaluated the proposed framework on a dataset of 80 video sequences consisting of a total of approximately 450 natural driving scenes of car crashes and abnormal driving events captured by traffic surveillance cameras. The video samples in the created dataset were collected from various publicly accessible websites and downloaded free of charge, which are representative of a wide range of road types including straight ramps, curved ramps, bridges, crossings, etc. They also involved various car events such as turning right, turning left, entering and leaving.

In order to quantitatively and comprehensively benchmark the performance of the proposed framework in predicting driving errors and their expected vehicular accidents, the receiver operating characteristic

Table 1. A brief performance comparison between our method and various related state-of-the-art methods.

Work	TP-r (%)	FP-r (%)
Proposed method	98.5	4.20
Chen et al. [20]	96.8	5.87
Ren et al. [37]	94.6	4.36
Arceda and Riveros [38]	89.0	6.2
Lu et al. [27]	87.8	7.33
Singh and Chalavadi [22]	77.5	2.25

(ROC) curve is used, as a measure of discrimination ability, to depict the trade-off between hit or true positive (TP) and false positive (FP) rates corresponding to a particular threshold determination and it plots a 2-D graph, with FP on x-axis and TP on y-axis. The values of TP and FP rates are calculated as:

$$TP-r = \frac{\text{correctly predicted accidents}}{\text{Total accidents}}$$

$$FP-r = \frac{\text{incorrectly predicted non-accidents}}{\text{Total non-accidents}}$$

The plot of the ROC curve of the presented system for vehicular accident prediction and detection is depicted in Fig. 6. As seen in the figure, the overall performance of the system yields mostly promising results, achieving a detection rate of 98.5% with a false alarm rate of 4.2%. Perhaps most importantly, overall, these results illustrate that the proposed method performs favorably against previous state-of-the-arts [20, 22, 27, 37, 38], in terms of both successful hit rate and low false alarm rate (see Table 1).

Sample snapshots from vehicular accidents correctly detected by the proposed system are shown in Fig. 7, where the system not only could potentially predict in advance the distracted driving behaviors, but also could uniquely identify, record and track the trajectory of each vehicle by which the accident was caused. In addition, the average time in predicting a vehicular accident is about 1.6523 sec.

All computational experiments were performed on a Dell Optiplex 9020 desktop PC with CPU Intel Core i7 4th Gen. i7-4770 3.4 GHz, 8 GB



Fig. 7. Sample snapshots from some video clips for vehicular accidents correctly detected by the presented system.

RAM, running Windows 10 Pro (64-bit). All of the algorithms and the various simulated parts of the proposed framework were implemented in C++ using Microsoft Visual Studio 2015 IDE and the open-source OpenCV library for image and video processing. The relatively low computational requirements allow the system to run in real-time (approximately at an average of 28 fps). This improved timing performance enables the proposed method to provide unique latency guarantees for real-time traffic surveillance and monitoring applications.

6. Conclusion

This paper has introduced a fuzzy vision-based framework for real-time vehicular accident prediction and detection, based on motion temporal template analysis and fuzzy time-slicing. Within the framework, an efficient spline neural network model is trained on a small and optimal set of potential local features, including moment invariants for detecting irregular vehicle behavioral patterns and thus predicting vehicular accidents just before they occur. Experimental results on realistic vehicular accident videos have shown the superiority of the proposed framework over recent state-of-the-art methods, in terms of detection rate and false-positive rate, while also still maintaining real-time performance. Future work will be directed toward further improvement of the predictive performance and robustness of the approach through using more video data with varying severity levels of crashes and supplementary methods (e.g., multi-view cameras and few-shot learning), so as ultimately to facilitate the identifying of various severity levels of vehicular accidents.

Declarations

Author contribution statement

Samy Bakheet: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Ayoub Al-Hamadi: Contributed reagents, materials, analysis tools or data.

Funding statement

Ayoub Al-Hamadi was supported by Bundesministerium für Bildung und Forschung (AutoKoWAT-3DMA Nr. 13N16336) and DFG-Project Nr. GZ: Al 638/15-1.

Data availability statement

No data was used for the research described in the article.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

References

- [1] World Health Organization, Violence, injury prevention, and world health organization, Global status report on road safety 2018: supporting a decade of action, World Health Organization, 2018.
- [2] S. Sadek, A. Al-Hamadi, B. Michaelis, U. Sayed, Real-time automatic traffic accident recognition using HOF, in: Proceedings of the 20th International Conference on Pattern Recognition (ICPR'10), Istanbul, Turkey, 2010, pp. 3348–3351.
- [3] R. Bergel-Hayat, M. Debarh, C. Antoniou, G. Yannis, Explaining the road accident risk: weather effects, *Accid. Anal. Prev.* 60 (2013) 456–465.
- [4] M.J. Cassidy, S.B. Anani, J.M. Haigwood, Study of freeway traffic near an off-ramp, *Transp. Res., Part A, Policy Pract.* 36 (6) (2002) 563–572.
- [5] Y.-H. Wen, T.-T. Lee, H.-J. Cho, Hybrid models toward traffic detector data treatment and data fusion, in: Proc. of IEEE Conf. on Networking, Sensing and Control, 2005, pp. 525–530.
- [6] J. Tamerius, X. Zhou, R. Mantilla, T. Greenfield-Huitt, Precipitation effects on motor vehicle crashes vary by space, time, and environmental conditions, *Weather Clim. Soc.* 8 (4) (2016) 399–407.
- [7] F.-H. Chan, Y.-T. Chen, Y. Xiang, M. Sun, Anticipating accidents in dashcam videos, in: S.-H. Lai, V. Lepetit, K. Nishino, Y. Sato (Eds.), *Computer Vision – ACCV 2016*, Springer International Publishing, Cham, 2017, pp. 136–153.
- [8] Y. Yao, M. Xu, Y. Wang, D.J. Crandall, E.M. Atkins, Unsupervised traffic accident detection in first-person videos, in: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 2019, pp. 273–280.
- [9] Zu hui, Xie yaohua, Ma lu, Fu Jiansheng, Vision-based real-time traffic accident detection, in: Proceeding of the 11th World Congress on Intelligent Control and Automation, 2014, pp. 1035–1038.
- [10] S. Sadek, A. Al-Hamadi, B. Michaelis, U. Sayed, A statistical framework for real-time traffic accident recognition, *J. Signal Inf. Process. (JSIP)* 1 (2010) 77–81.
- [11] H.H. Pour, F. Li, L. Wegmeth, C. Trense, R.J. Doniec, M. Grzegorzec, R. Wismüller, A machine learning framework for automated accident detection based on multimodal sensors in cars, *Sensors (Basel, Switzerland)* 22 (2022).
- [12] A. Jahangiri, H.A. Rakha, T.A. Dingus, Adopting machine learning methods to predict red-light running violations, in: 2015 IEEE 18th International Conference on Intelligent Transportation Systems, 2015, pp. 650–655.
- [13] E. Ohn-Bar, M.M. Trivedi, Beyond just keeping hands on the wheel: towards visual interpretation of driver hand motion patterns, in: 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), 2014, pp. 1245–1250.
- [14] M. Meier, Car Color and Logo Recognition, CSE 190 A Projects in Vision and Learning, University of California, 2006.
- [15] C. Hermes, C. Wohler, K. Schenk, F. Kummert, Long-term vehicle motion prediction, in: 2009 IEEE Intelligent Vehicles Symposium, 2009, pp. 652–657.
- [16] S. Sivaraman, B. Morris, M. Trivedi, Learning multi-lane trajectories using vehicle-based vision, in: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), 2011, pp. 2070–2076.
- [17] J. Wiest, M. Höffken, U. Kreßel, K. Dietmayer, Probabilistic trajectory prediction with Gaussian mixture models, in: 2012 IEEE Intelligent Vehicles Symposium, 2012, pp. 141–146.
- [18] S. Xia, J. Xiong, Y. Liu, G. Li, Vision-based traffic accident detection using matrix approximation, in: 2015 10th Asian Control Conference (ASCC), 2015, pp. 1–5.
- [19] H. Ullah, M. Ullah, H. Afridi, N. Conci, F.G.B. De Natale, Traffic accident detection through a hydrodynamic lens, in: 2015 IEEE International Conference on Image Processing (ICIP), 2015, pp. 2470–2474.
- [20] Y. Chen, Y. Yu, T. Li, A vision based traffic accident detection method using extreme learning machine, in: 2016 International Conference on Advanced Robotics and Mechatronics (ICARM), 2016, pp. 567–572.
- [21] B. Maaloul, A. Taleb-Ahmed, S. Niar, N. Harb, C. Valderrama, Adaptive video-based algorithm for accident detection on highways, in: 2017 12th IEEE International Symposium on Industrial Embedded Systems (SIES), 2017, pp. 1–6.
- [22] D. Singh, C.K. Mohan, Deep spatio-temporal representation for detection of road accidents using stacked autoencoder, *IEEE Trans. Intell. Transp. Syst.* 20 (3) (2019) 879–887.
- [23] W. Sultani, C. Chen, M. Shah, Real-world anomaly detection in surveillance videos, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 6479–6488.
- [24] T. Suzuki, H. Kataoka, Y. Aoki, Y. Satoh, Anticipating traffic accidents with adaptive loss and large-scale incident db, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 3521–3529.
- [25] H. Kataoka, T. Suzuki, S. Oikawa, Y. Matsui, Y. Satoh, Drive video analysis for the detection of traffic near-miss incidents, in: 2018 IEEE International Conference on Robotics and Automation (ICRA), 2018, pp. 3421–3428.
- [26] A.P. Shah, J. Lamare, T. Nguyen-Anh, A. Hauptmann, Cadp: a novel dataset for cctv traffic camera based accident analysis, in: 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2018, pp. 1–9.
- [27] Z. Lu, W. Zhou, S. Zhang, C. Wang, A new video-based crash detection method: balancing speed and accuracy using a feature fusion deep learning framework, *J. Adv. Transp.* (2020) 1–12.
- [28] A.F. Bobick, J.W. Davis, The recognition of human movement using temporal templates, *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (3) (2001) 257–267.
- [29] S. Bakheet, A. Al-Hamadi, M.A. Mofaddel, Recognition of human actions based on temporal motion templates, *Br. J. Appl. Sci. Technol.* 20 (5) (2017) 1–11.
- [30] J. Flusser, T. Suk, Pattern recognition by affine moment invariants, *Pattern Recognit.* 26 (1993) 167–174.
- [31] S. Sadek, A. Al-Hamadi, B. Michaelis, U. Sayed, Human activity recognition: a scheme using multiple cues, in: Proceedings of the International Symposium on Visual Computing (ISVC'10), Las Vegas, Nevada, USA, vol. 1, 2010, pp. 574–583.
- [32] S. Bakheet, A. Al-Hamadi, R. Youssef, A fingerprint-based verification framework using Harris and SURF feature detection algorithms, *Appl. Sci.* 12 (4) (2022).
- [33] S. Bakheet, A. Al-Hamadi, A hybrid cascade approach for human skin segmentation, *Br. J. Math. Comput. Sci.* 17 (6) (2016) 1–14.
- [34] S. Bakheet, A. Al-Hamadi, A framework for instantaneous driver drowsiness detection based on improved HOG features and naïve Bayesian classification, *Brain Sci.* 11 (2) (2021) 240–254.

- [35] S. Bakheet, A. Al-Hamadi, Computer-aided diagnosis of malignant melanoma using Gabor-based entropic features and multilevel neural networks, *Diagnostics* 10 (2020) 822–837.
- [36] E. Catmull, R. Rom, *Computer Aided Geometric Design, Chapter a Class of Local Interpolating Splines*, Academic Press, 1974, pp. 317–326.
- [37] J. Ren, Y. Chen, L. Xin, J. Shi, B. Li, Y. Liu, Detecting and positioning of traffic incidents via video-based analysis of traffic states in a road segment, *IET Intell. Transp. Syst.* 10 (6) (2016) 428–437.
- [38] V. Machaca Arceda, E. Laura Riveros, Fast car crash detection in video, in: 2018 XLIV Latin American Computer Conference (CLEI), 2018, pp. 632–637.