International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

Distinctly Trained Multi-Source CNN for Multi-Camera Based Vehicle Tracking System

Sanda Sri Harsha, Harika Simhadri, Karaganda Raghu, K.V. Prasad

Abstract-In the last few years the exponential rise in the demands or robust surveillance systems have revitalized academia-industries to achieve more efficient vision based computing systems. Vision based computing methods have been found potential for the different surveillance purposes such as Intelligent Transport System (ITS), civil surveillance, defense and other public-private establishment security. However, computational complexities turn-out to be more complicate for ITS under occlusion where multiple cameras could be synchronized together to track certain target vehicle. Classical texture, color based approaches are confined and often leads false positive outcome thus impacting decision making efficiency. Considering this as motivation, in this paper a highly robust and novel Distinctly Trained Multi-Source Convolutional Neural Network (DCNN) has been developed that exhibits pre-training of the real-time traffic videos from multiple cameras to track certain targeted vehicle. Our proposed DCNNvehicle tracking model encompasses multiple shared layers with multiple branchesof the source-specific layers. In other words, DCNN is implemented on each camera or source where it performs feature learning and enables a set of features shared by each camera, which is then learnt to identify Region of Interest (ROI) signifying the "targeted vehicle". Our proposed DCNNmodel trains each source input iteratively to achieve ROI representations in the shared layers. To perform tracking in a new sequence, DCNNforms a new network by combining the shared layers in the pre-trained DCNN with a new binary classification layer, which is updated online. This process enables online tracking by retrieving the ROI windows arbitrarily sampled near the previous ROI state. It helps achieving real-time vehicle tracking even under occlusion and dynamic background conditions.

Keywords: Multiple Camera based Vehicle Tracking, Vision Technology, , Convolutional Neural Network, Distinctly Trained Multi-Source CNN.

I. INTRODUCTION

In the last few years the exponentially rise vision technologies have broadened the horizon for superior surveillance systems with higher reliability and efficiency.

Revised Manuscript Received on 30 July 2019.

* Correspondence Author

Dr. Sanda Sri Harsha*, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur (Dist)- 522502, India.

Harika Simhadri, IT Consultant, Param Technologies Inc. Minneapolis, Dallas.

Dr. Katragadda Raghu, Department of Business Management, V R Siddhartha Engineering College, Vijaywada, India.

Dr. K.V. Prasad, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur (Dist)- 522502, India.

Undeniably, vision technologies, software systems, and more importantly low cost hardware availability has broadened the uses of computer vision based surveillance systems serving numerous purposes such as civil monitoring and surveillance, industrial monitoring and control, surveillance systems for defense establishments and numerous other public-private establishments, commercial entities etc. Computer Vision (CV) based approaches have gained widespread momentum across globe towards visual surveillance purposes which has revitalized academia-industries to achieve more efficient and reliable solution. Amongst common purposes, CV based ITS systems have gained widespread significance, which has attracted academia-industries to enable better efficiency; however classical approaches, especially those exploiting morphological information, color, texture etc are confined to provide optimal and reliable response [1].Single camera based surveillance systems are often confined due to lack of ability to track vehicle under occlusion conditions or changing background scenarios [1].To achieve a robust solution, efforts have been made to enhance processes like background subtraction, feature extraction, object detection and tracking, etc; however numerous real-time functions such as occlusion, path changes, transient and temporary exclusion from the view and drifting, similar vehicle features (i.e., color, texture, size etc) makes classical approaches confined [1-4]. In general, in the moving vehicle tracking process occlusion takes place once certain part of the concept region or vehicle ROI becomes out of vision or inaccessible by camera sensor to continue continuous tracking. Normally, occlusion occurs when a part of the moving vehicle gets covered due to other. Under such conditions tracking a vehicle becomes highly complicate and tedious task. Such gaps have motivated researchers to exploit multiple camera based vision computing methods for reliable vehicle tracking even under changing background and occlusion conditions [5][6]. Undeniably, vehicle detection in conjunction with inherent features like vehicle morphological size, shape, color, texture, stopped, and type etc can be significant for vehicle tracking in ITS decision systems [6-8]; however such approaches are confined and often exhibits false alarm during occlusion and changing background conditions [7][8]. In addition, feature similarity with neighbouring vehicle or the concept region can also affect overall accuracy of the tracking system. Recalling the fact that single camera based approaches are limited for moving vehicle tracking under occlusion, especially under long-term occlusion conditions and therefore multiple camera based systems have been recommended [9].

Multiple cameras based IT Smonitoring can employs multiple (synchronized) cameras placed at the different locations or orientation to detect and track continuously with reference to certain common features. However, performing so in real-time environment is highly tedious task, especially in terms of time and accuracy centric demands.

Moving vehicle tracking demands certain robust and more efficient systems which could be able to detect ROI within a very small span. In other words, the vision technique must be able to detect concept region or ROI transiently once it leaves vision rage of one camera and enters into the other. In some existing vehicle tracking systems, varied methods have tried to use depth information retrieved from multiple cameras to continue undisrupted vehicle tracking; however, to achieve robust solution demands fulfilling three key conditions including robustness, adaptivity and real-time processing [5]. Undeniably, numerous methods have been developed by exploiting efficacy of particle filters [26], three dimensional depth map projection [10]to perform moving object tracking under occlusion. The other approach which has gained wide-spread attention globally is CNN or deep learning concepts; however has been used majorly in the problem of image type classification and has been so far secluded for real-time vision based trackingpurposes. Undeniably, deep neural network methods have exhibited appreciable performance for image classification purposes; however its efficacy for multiple camera based vehicle tracking has not yet explored. CNN algorithms have played significant role for the different computer vision purposes like image classification in computer aided diagnosis (CAD) purposes. However, realizing the need of time and computationally efficient system authors developed recurrent CNN (RCNN) model [50] that employed low training data to achieve time efficiency. Unfortunately, RCNN required manipulatingdata to achieve expected performance which can be tedious task for moving vehicle tracking. To achieve a better solution, authors [Harsha paper] exploited transferable CNN technique to be used for vehicle classification. However, it could not perform vehicle tracking under occlusion. In addition, use of AlexNet for real-time moving video feature extraction and learning is tedious and hence as an optimal solution, in this paper we have developed a novel and robust Distinctly Trained Multi-Source CNN (DCNN) model for moving vehicle detection and tracking. The proposed system employs multiple DCNN model at each source (i.e., camera) that enables generation of a shared (feature) layer and performs binary classification at each source to detect and track ROI continuously. Some of the key novelties of the proposed multiple camera based vehicle tracking system are distinct source specific feature learning and ROI detection, shared layer feature learning, feature selection, N-fold cross validation, real-time ROI detection and tracking. The remaining sections of the presented manuscript are divided as follows: Section II discusses the related work, which is followed by the discussion of the problem formulation in Section III. Section IV presents the proposed system and its implementation, while the results obtained and its discussion is given in Section V. Overall research conclusion and future scopes are discussed in Section VI. References used are mentioned at the end of the manuscript.

II. RELATED WORK

This section discusses some of the key literatures pertaining to the vision based vehicle detection and tracking. Authors [11] developed a vision based vehicle detection model where they applied Gaussian Mixture Model (GMM) based background segmentation; Gabor based feature extraction followed by Hole Filling algorithm. To deal with dense vehicle classification, a vector sparse coding scheme with SVM was proposed in [12]. Applying sparse coding technique, authors projected features to the high dimensional vector which was learned and classified using Support Vector Machine (SVM) algorithm to perform classification. The combined shape and gradient feature based classification were proposed in [13][14], where authors applied vehicle shape to perform classification. Authors used silhouettes in the omnidirectional video frames to estimate shape information. For gradient based classifier Histogram of Oriented Gradients (HOG) features were obtained and the duo-features (shape and HOG) were used to perform classification. The features like geometry, number plate location and shape were used as input of dynamic Bayesian network (DBN) for vehicle detection and classification [15], where GMM was used to calculate probability distribution of features. Authors could not address the detection issues under varying illuminations, occlusion and frame dynamicity. A sparse learning based vehicle detection and classification model was proposed in [14]. Later, in [16] sparse coding and spatial pyramid matching scheme was used for vehicle detection and classification for ITS purposes, where they extracted the patch based sparse features using discriminate dictionary. Authors classified the extracted features using histogram intersection kernel based SVM classifier. In [17] multi-resolution vehicle recognition (MRVR) scheme was developed that in conjunction with cascade boosted classifiers performed vehicle detection and tracking. Authors used HAAR and HOG features to perform vehicle concept region detection and classification [18]. An enhanced approach with multi-feature fusion was developed in [19], where bothlocal as well as global featuresof the vehicle concept region were obtained using Spatial Invariant Feature Transform (SIFT). The extracted fetaures were processed for SVM based classification to achieve real-time vehicle detection and classification [19]. Considering deep neural network (DNN) based approaches, to achieve more efficient performance and reliability, authors [20][21] recommended using higher layer features. Author [20] used PHOG and LBP-EOH extracted using DNN for vehicle detection and classification; however could not perform tracking under dynamic background and occlusion conditions. In [22] authors used appearance features of the vehicle to perform detection and classification using semi-supervised CNN algorithm. In [23], shape based multi-class classification model was developed where the concavity property of vehicles like buses and sedans were used for detection classification; however it could not address the issue of tracking under occlusion or dynamic background conditions. A DBN was designed in [24] where HOG features and Eigen features were used together to perform vehicle detection and tracking.Unlike single base classifier based approaches, authors applied ensemble classification with SVM, K-NN, random forest and multiple-layer perceptrons (MLPs) algorithms.In [25], vehicle morphology was used where to deal with occlusionROI accumulative curve method and Fuzzy Constraints Satisfaction Propagation (FCSP) were developed. Retrieving the Time-Spatial Images (TSI) from the surveillance video, they eliminated shadowed region using SVM and Deterministic Non-Model Scheme (DNMS). Similarly, in [21], authors used conventional median filter and Otsu method-based background subtraction for vehicle detection; however was unable to detect moving vehicle under occlusion and changing background conditions. In [22] authors developed a vehicle detection, tracking, classification system; however could not address the tracking under occlusion condition. Considering multiple camera based vehicle tracking particle filters based approaches have also been explored [26]. To achieve vehicle tracking under occlusion, authors used blob-based ROI association and fragment based ROI construction schemes where features like color, gradient, were used (by fusing together).Authors used particle filtering [27] and sparse representation [28] to perform feature matching amongst multiple images to perform tracking; however it was developed for human tracking under occlusion. In [29] authors used random forest mining concept to performtracking under occlusion. In their model, they hypothesized that the association between target motion and its 3D data points can be significant to perform tracking of th3e concept region. To achieve moving object tracking under overlapping and non-overlapping conditionsparametric ellipsoid in a 3D model was developed in [30]. To deal with occlusion, authors applied SIR-Particle filter along with encoding (of) the visibility and persistence to the state vector [30]. Authors [31, 32] used colour and texture features for human trackingwhere particle filter was used to perform precise human tracking under occlusion. Authors used weighted color histogram (WCH) and cellular local binary pattern histograms to alleviate the issue of the partial occlusion. Multiple cameras were used in [33] to perform continuous objet tracking where views from each camera were mapped continuously to perform vehicle tracking. In [34] authors used multiple view features learning and merging approach to deal with occlusion. Similarly, in [35] authors extracted ROI feature points containing movement pattern and location which was processed further using particle filtering to perform target tracking. In [36], authors developed a fast moving object tracking scheme using multiple cameras, where relative positions of the objects were used to estimate the speed of the moving to assist occlusion avoidance. Region covariance matrixes were applied to form target appearance model which was further enhanced with homography method to estimate the interactions among multiple cameras for object tracking. Simlar to [26] and [34], authors [37, 38] suggested feature fusion from multiple cameras to perform tracking under occlusion. Multiple camera view-points with distinct features including appearance, depth, and occlusion were applied to perform tracking in [39]. In [50], authors applied SIFTbased multi-view, joint sparse pattern analysis model to perform continuous tracking however its efficiency for real-time moving vehicle tracking under occlusion remained unexplored. As an enhanced solution, authors [40, 41] developed a multi-view cooperative tracking system using SIFT features that further helped retrieving correlation in between the inter-frame and inter-view (multiple) camera views to perform tracking under changing background and occlusion.

III. PROBLEM FORMULATION

This is the matter of fact that occlusion in moving vehicles turns out the process of tracking highly complicate, especially under occlusion conditions.On contrary, alleviating this issue is of utmost significance for accurate vehicle surveillance and allied decision purposes. Considering about the existing approaches CNN has exhibited satisfactory performance in major computer vision applications [42-49] and hence to achieve more effective solution, authors have recommended optimizing CNN to perform feature extraction and learning over large scale data. Towards this objective Recurrent CNN (RCNN) has performed satisfactory [50], however it requires continuous data manipulation which seems difficult in real-time moving vehicle tracking under occlusion. It complexity can even be more complicate in case of multiple camera based approaches. Undeniably, unlike traditional vision based approaches CNN can be a better alternative due to its fine-grained feature extraction and learning ability that as a result could help achieving more reliable and accurate tracking of the concept region (i.e., vehicle). However, the allied computational complexity to learn over gigantic data with large number of features (especially in real-time computation) confines its efficacy to perform transient target detection and classification, which is must for vehicle tracking purposes. On the other hand, the use of multiple camera based models can be of utmost significance for vehicle tracking even under occlusion, provided each source or camera frame-sequence (data) is processed distinctly using a robust feature extraction and learning method and synchronized together to perform target (say, concept region or ROI) detection and tracking. However, maintaining optimal performance with higher accuracy and reliability too is inevitable. With this motive, this research intends to develop an efficient CNN model with low computational overheads or complexity to perform vehicle tracking over the (shared) features from each source dynamically. In this paper, effort has been made on developing an online-learning model by exploiting a pool of CNNs to be implemented at each source where it may perform binary classification for the occurrence of the concept region iteratively. Her, a novel CNN model named Distinctly Trained Multi-Source CNN (DCNN) has been developed that embodies efficacy of large scale multi-source visual tracking ability to perform pre-training of CNN so as to achieve ROI representation and tracking. The detailed discussion of the proposed DCNN model is given in Fig. 1. DCNN learns the shared ROI representations from multiple camera and respective distinct video sequences (individually) to perform ROI detection and continuous tracking. In our proposed model each video sequence is considered as a distinct source. Noticeably, we have considered video sequence from the three distinct sources or cameras, where on each source DCNN has been implemented.

DCNNemploysdifferent (distinct) branches of source-specific layers (it states that the DCNN is implemented on each source) to perform binary classification at the end of the network where it shares the common key features retrieved from all image sequences in the preceding layers for the representation learning.

In our proposed vehicle tracking model, each source input is trained distinctly and iteratively where the shared layers (say, shared key points or image information) are updated simultaneously. Applying this approach, the source-independent vehicle features or the information from the source-specific one can be obtained that can help for generic feature representation learning for ROI tracking. DCNN applies both lower layer as well as higher layer features to achieve high accurate performance. Learning the extracted ROI features vehicle tracking has been performed by DCNN. With a provided test data sequence, in DCNN model all branches of the binary classification layers used for training are removed and updated with a new single branch to identify ROI's total score matrix in the video or test sequence. Thus, the newly updated classification layer and the fully connected (FC) layers within the shared layers are followed by online fine-tuning when exhibiting tracking to adapt another (new) source data. Such update helps to deal with occlusion conditions. Additionally, we used a novel hard negative mining model during feature learning that enabled accurate concept region detection and tracking and thus achieved optimal performance for the vehicle detection under occlusion. The overall contributions made in this research can be stated as follows:

- A novel distinctly trained multi-source CNN (DCNN) model to perform distinctive source independent information retrieval from the source-specific one and retrieve shared ROI representations to perform vehicle tracking under occlusion.
- A novel multiple source (i.e., multiple camera) training model and feature fusion for vehicle tracking under occlusion.
- Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA) assisted feature selection for swift computation.
- Machine learning assisted classification for accurate vehicle detection and tracking over input video sequences (i.e., frames).

DCNN assisted vehicle tracking model in which CNN is (pre)trained by multiple source learning and iterative update with respect to the new sequence for learning source-specific information dynamically and/or adaptively.

IV. SYSTEM MODEL

This section of the research work presents the development and implementation of the proposed DCNN model for vehicle detection and tracking under occlusion conditions.

A. Multi-Source Learning (MSL)

The proposed DCNN model is one of the types of Multi-Domain Learning (MDL) methods [26][27][35] that signifies to the learning paradigm in which the training sequences or data are produced from various sources, such as multiple surveillance cameras, and where the related source information is used to facilitate distinctive learning for building a generic ROI representation. Though in the applications such as spam filtering, sentiment analysis etc the MDL approaches have been implemented; however its effectiveness and efficacy is not yet explored fully in the computer vision technologies. Amongst the numerous researches with MDL models, some of the key significant ones are concept region detection in video by using domain-weighted combination of the SVM [28], mixture transformation model for object classification [51] etc. It has been revealed from these research works that MDL is of paramount significance for detecting object with efficacy even under non-uniform conception region characteristics. Considering this as a motivation, here a robust MDL–based multiple source learning (MSL) scheme has been developed. This scheme aims at achieving independent sequence learning from the different sources and updates a new sequence with a common ROI representation in order to facilitate efficient tracking. Through this method the possibilities of detecting concept region and performing tracking can be achieved successfully even under the short-term and long term variations in ROI characteristics and other background conditions such as occlusion.The deep learning concept discussed above is implemented to efficiently perform tracking under occlusion conditions.

B. Distinctly Trained Multi-Source CNN Model (DCNN)

In this section of the paper, a detailed discussion of the DCNN model and its architecture has been presented in order to demonstrate desired vehicle tracking even under occlusion using inputs from the multiple cameras.

a) DCNN Architecture

A snippet of the proposed DCNN model is given as follows:

International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

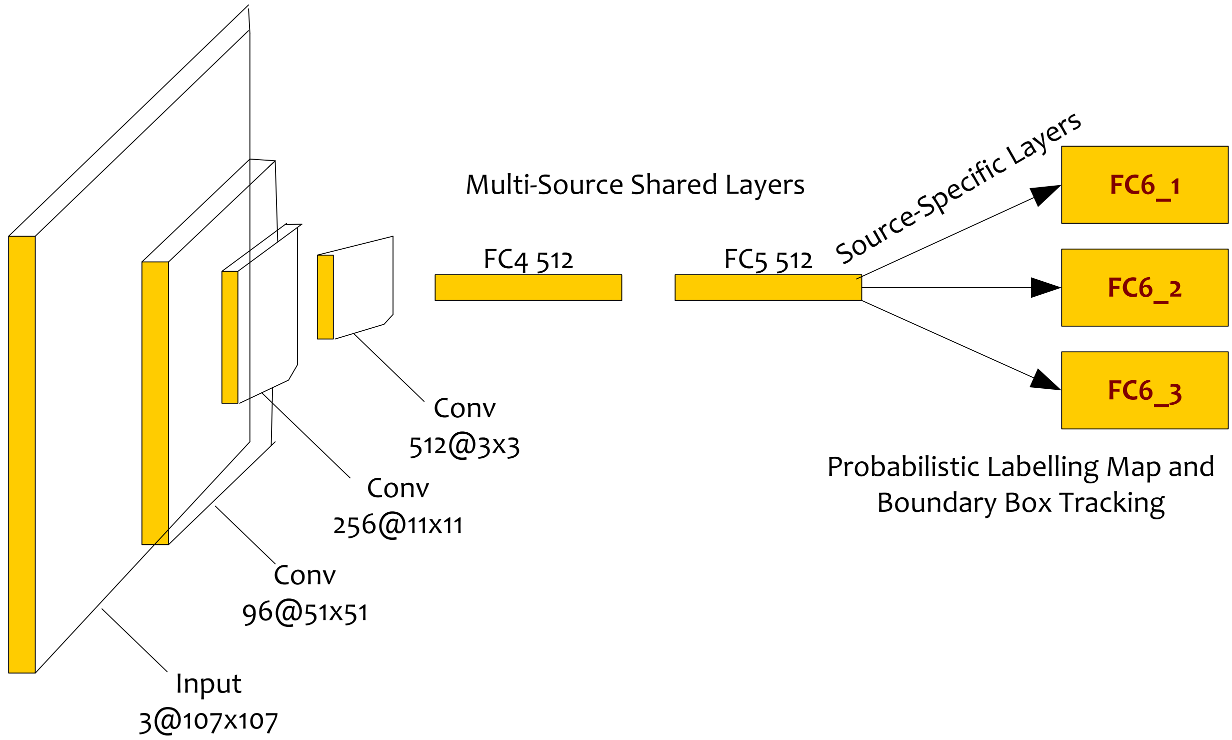

Fig. 1The architecture of DCNN based vehicle detection and tracking

The deep learning model presented in the figure (Fig. 1) clearly illustrates that initially input video stream and the consequent image frames with 107×107 RGB are given as input. There are five hidden layers in the proposed network model; three consecutive convolutional layers, labeled as conv 1-3 and two sequentially arranged Fully connected layers, labeled as FC 4-5. Additionally, the DCNN model has K unique branches and respectively K unique training sequences for the last FC layers, i.e. FC61-FC6K.Also there exists a similarity between the convolutional layers and the associated sections of the VGG-M network [52][53]. It is noteworthy that there is no similarity between the sizes of the feature map, since they are adjusted on the basis of the input size. The next two FC layers have 512 output units and are connected with the ReLUs and dropouts. Here, K branches denoting the distinct learning sequence have an individual binary classification layer with softmax cross-entropy loss. This binary classification layer facilitates the separation of vehicle region and the occlusion component. It is noteworthy that in the proposed model it has been assumed that all the inputs are presenting the sequence from identical orientation and are synchronized. The fc61-fc6K layers present in our DCNN model are called source-specific layers whereas all the preceding layers are called as shared layers. Noticeably the model is also lightweight in comparison to the classical models that are used in recognition functions like VGG-Nets [52][53]. This further adds-on to the superiority of our model and makes it suitable for vehicle tracking even under dynamic background changes and occlusion presence with efficacy. Though there are numerous reasons for the efficacy of DCNN, some significant ones are described here. The motive of visual tracking functions is to separate only two classes, i.e. vehicle region and occlusion component and thus the proposed model is less complex contrary to the visual recognition models such as ImageNet which functions with 1000 classes. Similarly, to efficiently perform vehicle tracking even under occlusion, a deep CNN becomes confined to exhibit accurate target identification as the spatial information often gets diluted as a network goes deeper [54]. These issues can also be alleviated with the proposed lightweight DCNN model.Since, vehicle tracking is based on the long distance, DCNN makes vehicle region of ROI, which is generally small and thus allows small input size which consequently minimizes the depth of the network and even reduces computational complexity. Predominantly, the small size network facilitates efficient visual tracking, especially when real-time training and testing are performed. It backs up the suitability of the proposed DCNN model to be used for real-time vehicle tracking under occlusion.

b) DCNN Learning Model

The proposed DCNN model is developed with a motive to perform training over a multi-source CNN disambiguating target, which may undergo occlusion condition and arbitrarily changes in the background conditions. Moreover, the surveillance data obtained from different domains can have distinct target's notion, background and occluding conditions. It is also noteworthy that there can be some generic properties or features in the data obtained from different sources (camera data) which can be utilized to perform ROI vehicle region representation in all sources. Some of these generic features are dimensions of vehicles (such as height and width) textures, motion blur, robustness of texture and illumination changes, scale variations etc. In order to extract the significant features, satisfying the generic features mentioned above across the multiple channels or sources, we distinguish source-independent information from source-specific one by employing our proposedMDL or MSL learning model. Undeniably, here the proposed DCNN is trained throughStochastic Gradient Descent (SGD) algorithm, which gathers source information from the individual cameras and handles it exclusively in iterations. Here, in the kth iteration, DCNN is updated on the basis of a mini-batch that comprises the training samples from the ($k \bmod K$)th data sequence. In our proposed model, a single branch FC6 (kmodK) is used to allow data learning from the individual sources of data. The process keeps on running until the developed network coincides or the total number of iterations is completed. As in the proposed learning model, source-independent information is transferred or modeled into the shared layers or a single layer output (FC6/FC7) from which the significant generic feature representations or concept regions are retrieved.

c) DCNN assisted Vehicle Tracking under Occlusion

Previously, it has been discussed that when multi-domain learning is performed, the several branches of the sources-specific layers i.e., FC61-FC6K are replaced with a single branch FC6 which finally leads to the generation of a new test sequence. Then the fine-tuning of the newly produced source-specific layer is done and the FC layers are obtained in the share network simultaneously. After obtaining the single branch layer data the tracking function is executed. A snippet of the proposed tracking model is given as follows:

d) Tracking Control and DCNN Update

In order to achieve vehicle tracking under occlusion conditions with efficacy, here two supplementary attributes, i.e., robustness and adaptiveness have been considered, by long term as well as short term updates. While the long-term updates are performed at predefined regular intervals by utilizing the positive samples collected since a long period, the short-term updates are performed only when there is significant tracking failures due to occlusion or when the positive score of the detected and tracked ROI (i.e. vehicle concept region) is less than 0.5.This is performed by applying the positive samples in a short-term.In our proposed model, under above stated conditions DCNN applies the negative samples that have come across in the short-term, since the old negative samples might not be relevant to the current frame or could even be redundant. It is well known that including the old negative samples can cause inaccuracy in detecting the concept region or vehicle region and hence consequently lead to wrong prediction which will further cause poor tracking results.Hence, it can be affirmed that by employing the positive samples in a short-term can facilitate online and accurate tracking. In the proposed model a single network is used during vehicle tracking such that the updates can be made based on the needs like how quickly the region of vehicle changes. In order to know about the vehicle region state in each frame, sampling of the $\mathcal{N}$ target candidates $\mathcal{X}^1 \dots, \mathcal{X}^{\mathcal{N}}$ around the previous target state has been performed and thus the network yields positive scores $\mathrm{f}^+(\mathcal{X}^i)$ as well as negative scores $\mathrm{f}^-(\mathcal{X}^i)$ for each frame.

Here, the optimal target state $\mathcal{X}^*$ is estimated by identifying the ROI or concept region with the highest positive score. Mathematically,

$$\mathrm{x}^* = arg \max_{\mathrm{x}^i} \mathrm{f}^+(\mathcal{X}^i) \tag{1}$$

Since the proposed learning DCNNmodel also implements mini-batch mining concept, a brief of it is presented in the next section.

e) Mini-batch Mining

Pertaining to the real-time application of the vehicle tracking under occlusion such as traffic data, most of the negative instances or examples are generally trivial to perform tracking-by-detection paradigms, while merely a few off-putting negative examples can be significant to train a classifier. Hence, the classical classifiers such as SGD which considers training instances or samples for learning suffer from drifting issue because the occluders are not addressed properly. Hard negative mining [58] is one of the most appropriate solutions to overcome such issues as they allow switching between training and testing processes which further enables the identification of the hard negative samples, usually the false positives. Considering this as a motivation in this research work the concept of N-fold cross validation has been implemented since it ensures better performance for online learning and efficient vehicle tracking. In the proposed work the DCNN model has been integrated with mini-batch selection where in each and every iteration during learning, a mini-batch provides $\mathcal{M}^+$ positives and $\mathcal{M}_h{}^-$ hard negatives. Here, the detection of $\mathcal{M}_h{}^-$ examples is done by performing testing negative samples and then selecting the ones having highest $\mathcal{M}_h{}^-$ scores. The mini-batch based classification becomes more intricate when the execution and continuation of the learning phase of DCNN network turns out to be more discriminative.Hence, in this case reducing feature samples by performing optimal feature selection can be of utmost significance. With this motive, we have applied PCA and LDA feature selection and/or dimensional reductionmethod. A snippet of these algorithms is given as follows:

f) Principle Component Analysis(PCA)

Undeniably, achieve PCA intends to maintain low size of the feature samples which used to be high dimension; while maintaining most significant features (i.e., mini-batch positive samples).In general, the feature components extracted from PCA algorithm used to be the most expressive features (MEF), while LDA employs the most discriminating features (MDF) function. In PCA based approach distinct principle component (PCS) is generated for individual class. However, despite of retrieving the distance from the average principal component of each class, we have trained the PCA vectors using SVM. Here, we have applied radial basis function (RBF) kernel for SVM training. We have trained SVM to retrieve the largest feasible classification margin that signifies the lowest value of w in (2).

$$\frac{1}{2} w^T w + \mathrm{E} \sum \varepsilon_i \tag{2}$$

where $\varepsilon_i \geq 0$ and E is the error tolerance level.

In our proposed model, the training vectors have been categorized in labeled pairs $\mathcal{L}_i\ (x_i, y_i)$ where x_i states the training vector, while the class label of x_i is given by $y_i \in \{-1, 1\}$.

In classification, the hyperplane groups highest feasible points of the same class on the same side, while increasing the distance of either class from it. To achieve optimal classification accuracy we have applied 10-fold cross validation scheme. To perform testing, a test image data has been processed for PCS estimation which has been followed by its principle component classification using trained SVM.In DCNN, the p-dimensional feature spaces are turned to a new position called as principal axes in such a way that the variance of principal axis 1 is the highest, in succession by axis 2 and so on. There are some features of the targets in the vehicle detection and tracking that are highly correlated and are called homogeneous features. These feature attributes can be the most expressive features (MEF) to represent concept region or the target. To achieve it, PCA transfers feature attributes from different source-information to another new feature vector with no correlation among the elements. Hence, applying the PCA approach causes a significant reduction in feature elements which makes the computation more efficient and the finally retrieved feature vector is then projected for classification.The proposed DCNN model utilizes the obtained number of samples, i.e. 50% of the total inputs. Additionally, the model also performs the identification of negative samples without imposing additional false positive detector as usually done in the hard negative mining.

g) Linear Discriminant Analysis(LDA)

As discussed previously, PCA based schemes employ MEFs to perform classification. However, MEFs can't be the MDFs all the time. On the contrary, LDA can perform automatic feature selection that can enable efficient feature space for further classification. To alleviate the issue of high dimensionality, LDA has been initiated by employing PCA, where all the vehicle region data or ROI irrespective of the class label has been projected onto a single PCS. The dimension of the PCS has been confined by the total training image minus the number of classes. In our model, two distinct metrics have been estimated, intra-class scatter matrix $\mathbb{I}_{ICW}$and inter-class scatter matrix$\mathbb{I}_{IOS}$. Mathematically these matrixes have been estimated using (3) and (4).

$$\mathbb{I}_{ICW} = \sum_{i=1}^{C} \sum_{j=1}^{\mathcal{M}_i} (y_j - \mu_i)(y_j - \mu_i)^T \quad (3)$$

$$\mathbb{I}_{IOS} = \sum_{i=1}^{C} (\mu_i - \mu)(\mu_i - \mu)^T \quad (4)$$

In above equation, Csignifies the total number of classes, μ_i states the average vector of a class i, and $\mathcal{M}_i$ signifies the number of samples within i. Thus, the average of the average vectors is obtained as (5).

$$\mu = \frac{1}{C} \sum_{i=1}^{C} \mu_i \quad (5)$$

LDA method focuses on maximizing the inter-class scatter while reducing the intra-class scatter by increasing the ratio $det|\mathcal{S}_B|/det|\mathcal{S}_W|$. The significance of applying this ratio is that in case of non-singular $\mathbb{I}_{IOS}$ matrix, the ratio can be increased when the column vectors of the projection matrix $\mathcal{W}$ can be the eigenvectors of$\mathbb{I}_{ICW}{}^{-1}\mathbb{I}_{IOS}$. Here, the projection matrix $\mathcal{W}$ with $C-1$dimension assigns the training data onto a new space, usually called fisher vector. Thus, $\mathcal{W}$is applied for projecting all the training samples onto the fisher vector. The retrieved feature vector $\mathcal{F}_{VR} = (f_{1R}, f_{2R}, f_{3R}, \dots, f_{4096R})$ has been further used for classification using SVM classifier with polynomial kernel function.

h) Bounding Box Regression (BBR)

It has been revealed through some literatures that the approaches based on bounding box regression are generally efficient in object tracking and better localization accuracy; while the same is not true for classical methods, especially when there are occlusion conditions. Hence to alleviate such issues bounding box generation model has been developed and implemented. Due to the high level abstraction of the DCNN based features;here a data augmentation process that samples multiple positive instances around the target is implemented. Additionally, it cannot be ignored that DCNN faces adversaries in estimating tight bounding boxes surrounding the vehicle region; hence to deal with these adversaries our proposed model employs an enhanced regression assisted bounding box generation model called BBR model. In DCNN with the first frame of a multi-source frame sequence, we execute a simple linear regression method that exhibits the accurate target localization using Conv3 features of the samples near the target location. Similarly, in the ensuing video frames, the concept region or the target (here, vehicle) locations are estimated using above derived function (Eq. (1)). Thus, the precise target is localized using the regression concept if obtained targets are reliable (i.e., $f^+(x^*)$). In our proposed model, the BBR model is trained only in the first frame as it becomes complicate and time consuming for online update and even the incremental learning of the regression might not be more significant.

V. RESULTS AND DISCUSSION

In this paper an enhanced CNN model was developed to perform vehicle detection and tracking under occlusion. The developed DCNN model was applied for multiple camera based vehicle detection and tracking purposes where it was applied at each camera to perform feature extraction that eventually helps deriving a shared feature for further classification. DCNN was implemented at each source whereit performed feature extraction and learning simultaneously to perform concept region tracking over moving data sequence (video data from camera). The proposed model at first performed feature extraction followed by feature selection using PCA and LDA algorithms. The selected features were concatenated and mapped to a single feature model projected and retrieved at the FC6 layer of the proposed DCNN model. The retrieved features were learnt with DCNN modal that enabled binary classification signifying the presence of the concept region in consecutive frames. DCNN model comprised 5 convolutional layer and 2 fully connected layers, FC6 and FC7. Here, we performed classification at the FC6 layers that comprised of FC6_1 to FC6_3 features obtained from the three different sources
(Fig. 1),

which are then concatenated and mapped as final feature for which it performs two class classification using SVM. To simulate CNN learning rate has been assigned as 0.0001, while the drop out ratio is assigned a 0.5. To simulate the proposed model, road-traffic video data has been collected while assuring that it embodies occlusion condition during movement.Here we have synthesized own data collected from three distinct camera installed at three different locations on highway. Here, we considered each camera input as individual source and executed DCNN on each sources distinctly to perform (individual) feature extraction. As stated, the extracted features from each source are concatenated to form a single feature structure also called shared features for vehicle concept region or ROI tracking. Here, the overall traffic videos were split into 2100 frame or image sequences (700 frames from each video sequence). In our proposed model, each image sequence was preprocessed and was resized to 256×256 pixel dimensions. DCNN was completely developed using MATLAB 2017a where MatConvNet[57] tool was used to simulate on Intel i3 processor with 8GB memory. Noticeably, the overall proposed model exhibited vehicle tracking in three sequential phases. A snippet of these approaches is given as follows:

a). ROI or the concept region identification

To generate or retrieve ROI (also called concept region) in each video frame, we drawn $(N = 2100)$video frames in translational scale dimensions, $X_t^i = (x_t^i, y_t^i, S_t^i), i = 1, \dots, N$ from Gaussian distribution whose mean is the recent (say, lastly occurred) concept region state, (i.e.,X_{t-1}^*) so that the covariance is always $\mathrm{diag}(0.09r^2, 0.09r^2, 0.25)$. Here, the variable r states for the average of the dimensional constructs (i.e., width and height) of the vehicle in the previous frame. For ROI visualization bounding box was generated by estimating a $1.05S_i$ of the initial target scale.

b). Multi-Source (Cameras)Feature Training

To perform multi-camera (or multi-source) learning, a sample of 200 frames with equal proportion from each data source were taken into consideration where the overlap ratio with ground truth bounding boxes for positive and negative examples were higher than 0.7 and lower than 0.5. In this work, to perform bounding box regression a total of 400 training examples with different spatio-temporal parameters [12] were taken into consideration.

c). DCNN Learning

To perform learning with three cameras, DCNN was applied at each source where the learning rate of 0.0001 was assigned at the convolutional layers, while the same was assigned as 0.001 for FC layers. To initiate learning, at the initial frame the FC layers were trained for 10 iterations with learning rate of 0.0001 for FC4 and FC-5, while 0.001 for FC6.

For concept region representation and (dynamic ROI) update at each source FC layers were trained for 10 iterations with increased learning rate that enabled adaptation with the descendent frames or the detected concept region. To achieve computational efficiency we maintained weight decay and momentum as fixed value, i.e., 0.0005 and 0.9, respectively. To achieve comparative performance assessment, we used SIFT features as well for classification where SVM with polynomial kernel function was used to perform two class classification. Noticeably, the two-class classification states whether the ROI is available in current frame of not. In case of identifying ROI presence or the concept region in frame bounding box were generated for the target vehicle. In this paper to achieve reliable outcome, 10-fold cross validation method was applied. To assess performance analysis, we estimated confusion matrix that helped achieving Accuracy, Precision, F-Measure, Recall etc. A snippet of the definitions of these performance parameters is given as follows:

To assess accuracy of the developed ensemble classification model we have obtained confusion matrix characterizing true positive (TP), true negative (TN), false positive (FP) and false negative (FN). Retrieving these matrix values for each base learners as well as ensemble classifier, the performance has been obtained in terms of accuracy, precision, f-measure, recall and specificity. The definitions of these performance variables are given in Table I.

TABLE I. PERFORMANCE PARAMETERS

Parameter	Mathematical Expression	Definition
Accuracy	$\frac{(TN+TP)}{(TN+FN+FP+TP)}$	Signifies the proportion of predicted fault prone modules that are inspected out of all modules.
Precision	$\frac{TP}{(TP+FP)}$	States the degree to which the repeated measurements under unchanged conditions show the same results.
F-measure	$2.\frac{Recall.Precision}{Recall+Precision}$	It combines the precision and recall numeric value to give a single score, which is defined as the harmonic mean of the recall and precision.
Recall	$\frac{TP}{(TP+FN)}$	It indicates how many of the relevant items are to be identified.

The results obtained for the proposed DTNN and SIFT based vehicle tracking system are given in Table II.

TABLE II. PERFORMANCE OBTAINED

Parameter	DCNN PCA (%)	SIFT PCA (%)	DCNN LDA (%)	SIFT LDA (%)
Accuracy	87.96	84.98	97.01	89.35
Precision	82.91	83.37	92.18	92.5

Recall	83.62	85.41	92.38	90.31
F-Measure	82.64	84.37	92.26	90.4

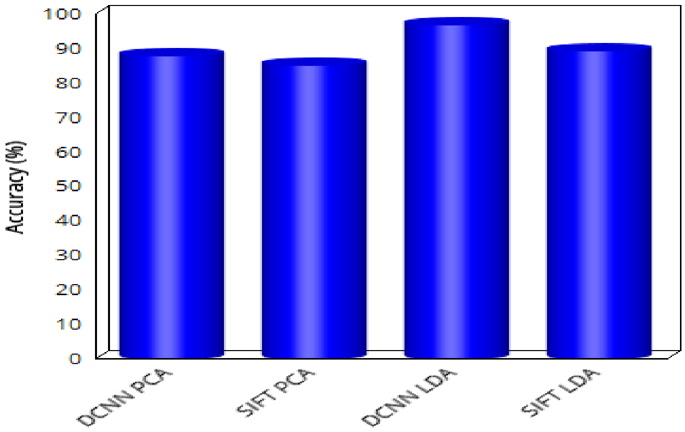

Fig. 2 Accuracy assessment

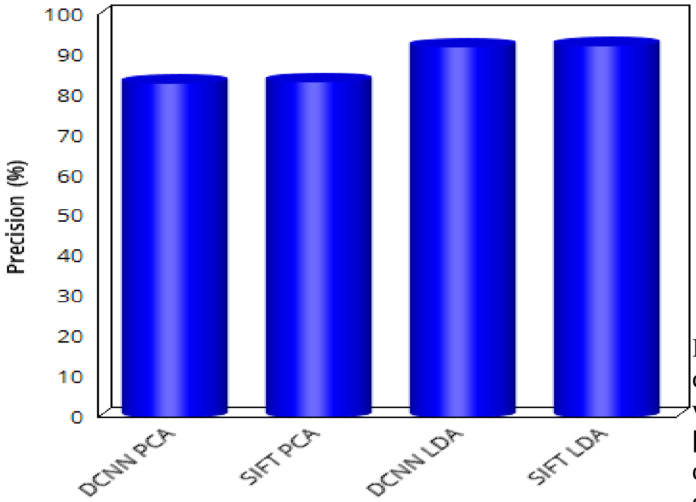

Fig. 3 Precision assessment

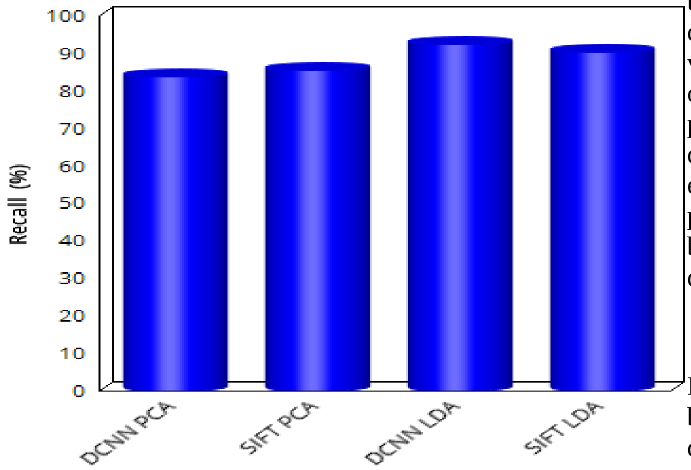

Fig. 4 Recall assessment

Fig. 2 presents the accuracy of the proposed DCNN based vehicle tracking system where it can be visualized that the DCNN with LDA feature selection outperforms other approaches. Considering existing SIFT based model, DCNN-LDA exhibits better accuracy (97.01%). On contrary, SIFT with PCA and LDA selected features exhibit accuracy of 84.98% and 89.35%, respectively. Similarly, DCNN-LDA retains better precision (92.18%) than other approaches (Fig. 3). Recall performance too reveals that the DCNN-LDA model outperforms other approaches (Figure 4). Fig. 5 presents the F-Measure performance where it can be observed that the DCNN-LDA shows highest value (92.26%) followed by SIFT-LDA (90.4%). The overall performance reveals that the DCNN_LDA can be a potential approach to perform online vehicle detection and tracking under occlusion conditions.

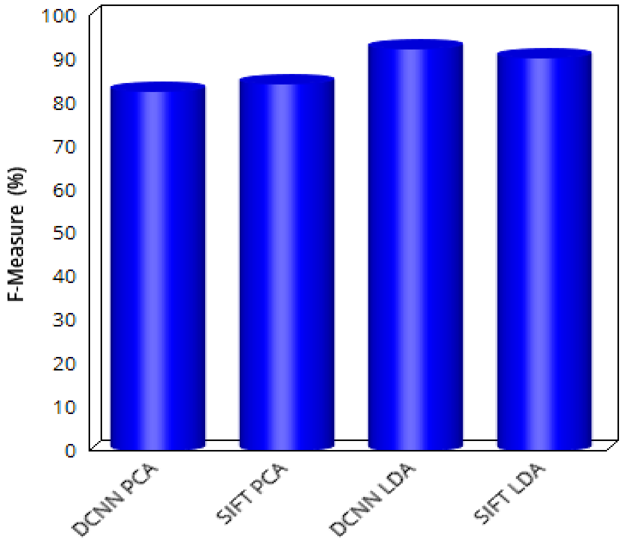

Fig. 4 F-Measure assessment

In addition to the above discussed results, we have compared the performance of the proposed DCNN model; we considered different existing approaches such as [56][57]. Authors in [56] applied self learning based vehicle detection and tracking (SLTD) where the average accuracy obtained was 94.86%, which is lower than our proposed DCNN-LDA model that exhibits 97.01% detection and tracking accuracy. Similarly, authors in [57] had developed deep representation and semi-supervised learning based vehicle detection and tracking model could achieve accuracy of 92.1%. Observing these results, it is affirmed that the proposed model outperforms other approaches for vehicle detection and tracking while assuring optimal performance even under occlusion conditions. Thus, observing overall performance it can be stated that the proposed approach can be of utmost significance for online vehicle tracking under dynamic background and occlusion scenarios.

VI. CONCLUSION

In this paper a highly robust and efficient multiple cameras based vehicle tracking system under occlusion has been developed. Unlike classical visual tracking system, the proposed Distinctly Trained Multi-Source CNN Model (DCNN) performs pre-training over a large datasets originating or coming from multiple cameras installed at the highway with different views or orientation. Unlike classical vehicle tracking approaches DCNN model performs independent feature learning over the video frame coming from multiple cameras and performs leaning to identify ROI or vehicle region in the frames.

Here, the use of shared layers enabled low dimensional features and allied computation that made DCNN to exhibit swift performance overall real-time vehicle tracking under occlusion. In addition to the shared layer, the proposed model incorporated multiple branches of source-specific layers that assisted concept region or Region of Interest Tracking with multiple cameras views. Since, DCNNexhibits training distinctly over the frame sequences it enabled each branch to perform support vector machine based binary classification for vehicle region or ROI tracking. The cumulative ROI and occlusion region identification in conjunction with bounding box generation enabled DCNNto exhibit vehicle region identification and tracking under occlusion.To perform vehicle tracking in a new sequence, DCNNforms a new network by combining the shared layers with a new binary classification layer (i.e., Fully Connected Layers), which is updated online. This overall process enabled our model to exhibit robust tracking under occlusion conditions.

REFERENCES

1. Goo, J., et al., 2008, "Tracking and Segmentation of Highway Vehicles in Cluttered and Crowded Scenes," in Applications of Computer Vision, IEEE Workshop, pp. 1-6.
2. Vasu, L., 2010, "An effective step to real-time implementation of accident detection system using image processing," Master of Science, Oklahoma State University, USA.
3. Jie, Z., et al., 2007, "Moving Vehicle Detection for Automatic Traffic Monitoring," Vehicular Technology, IEEE Transactions, 56, pp. 51-59.
4. Iwasaki, Y., and Itoyama, H., 2006, "Real-time Vehicle Detection Using Information of Shadows Underneath Vehicles," in Advances in Computer, Information, and Systems Sciences, and Engineering, K. Elleithy, et al., Eds., ed: Springer Netherlands, pp. 94-98.
5. Wei, Z., et al., 2008, "Multilevel Framework to Detect and Handle Vehicle Occlusion," Intelligent Transportation Systems, IEEE Transactions, 9, pp. 161-174.
6. Huang, D. Y. et al., 2012, "Reliable moving vehicle detection based on the filtering of swinging tree leaves and raindrops," J. Vis. Comun. Image Represent, 23, pp. 648-664.
7. R. Feris, A. Datta, S. Pankanti and M. T. Sun, "Boosting object detection performance in crowded surveillance videos," Applications of Computer Vision (WACV), 2013 IEEE Workshop on, Tampa, FL, 2013, pp. 427-432.
8. C. T. Chu, J. N. Hwang, K. M. Lan and S. Z. Wang, "Tracking across multiple cameras with overlapping views based on brightness and tangent transfer functions," Distributed Smart Cameras (ICDSC), 2011 Fifth ACM/IEEE International Conf. on, Ghent, 2011, pp. 1-6.
9. J. Harguess, C. Hu, and J. K. Aggarwal, "Occlusion robust multi-camera face tracking," CVPR 2011 Workshop, Colorado Springs, CO, 2011, pp. 31-38.
10. F. Dionnet, and E. Marchand, "Robust stereo tracking for space applications," 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems, San Diego, CA, 2007, pp. 3373-3378.
11. A. Nurhadiyatna, *et al.*, "Gabor filtering for feature extraction in real time vehicle classification system," *9th International Symposium on Image and Signal Processing and Analysis (ISPA), Zagreb*, pp. 19-24, 2015.
12. Ma, *et al.*, "Urban vehicle classification based on linear SVM with efficient vector sparse coding," *Information and Automation (ICIA), IEEE International Conference on, Yinchuan*, pp. 527-532, 2013.
13. H. C. Karaimer, *et al.*, "Combining Shape-Based and Gradient-Based Classifiers for Vehicle Classification," *2015 IEEE 18th International Conf. on Intelligent Transportation Systems, Las Palmas*, pp. 800-805, 2015.
14. X. Chen, *et al.*, "Vehicle representation and classification of surveillance video based on sparse learning," *in China Communications*, vol. 11, no. 13, pp. 135-141, 2014.
15. Y. Liu and K. Wang, "Vehicle classification system based on dynamic Bayesian network," *Service Operations and Logistics, and Informatics (SOLI), 2014 IEEE International Conference on, Qingdao*, pp. 22-26, 2014.
16. Y. Peng, *et al.*, "Vehicle classification using sparse coding and spatial pyramid matching," *17th International IEEE Conference on Intelligent Transportation Systems (ITSC), Qingdao*, pp. 259-263, 2014.
17. B. Zhang, *et al.*, "Vehicle Classification with Confidence by Classified Vector Quantization," *in IEEE Intelligent Transportation Systems Magazine*, vol. 5, no. 3, pp. 8-20, 2013.
18. Y. Hu et al., "Algorithm for vision-based vehicle detection and classification," *2013 IEEE International Conference on Robotics and Biomimetics (ROBIO), Shenzhen*, pp. 568-572, 2013.
19. W. Ma, *et al.*, "Vehicle classification based on multi-feature fusion," *Wireless, Mobile and Multimedia Networks (ICWMMN 2013), 5th IET International Conference on, Beijing*, pp. 215-219, 2013.
20. H. Qian, *et al.*, "Vehicle classification based on the fusion of deep network features and traditional features," *Advanced Computational Intelligence (ICACI), 2015 Seventh International Conf. on, Wuyi*, pp. 257-262, 2015.
21. Y. Zhou, *et al.*, "Vehicle Classification using Transferable Deep Neural Network Features," *arXiv:1601.01145v1 [cs.CV]* 6 Jan 2016.
22. [22] Z. Dong, *et al.*, "Vehicle Type Classification Using Unsupervised Convolutional Neural Network," *Pattern Recognition (ICPR), 2014, 22nd International Conference on, Stockholm*, pp. 172-177, 2014.
23. H. Gu, *et al.*, "Vehicle size classification for real time intelligent transportation system," *Conference Anthology, IEEE, China*, pp. 1-5, 2013.
24. B. Zhang, "Reliable Classification of Vehicle Types Based on Cascade Classifier Ensembles," *in IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 322-332, March 2013.
25. C. Y. Chen, *et al.*, "Vehicle classification and counting system," *Audio, Language and Image Processing (ICALIP), 2014 International Conference on, Shanghai*, pp. 485-490, 2014.
26. C. Nigam, R. V. Babu, S. K. Raja, and K. R. Ramakrishnan, "Feature Fusion for Robust Object Tracking Using Fragmented Particles," *2007 First ACM/IEEE International Conference on Distributed Smart Cameras*, Vienna, 2007, pp. 283-290.
27. T. Zhang, B. Ghanem, and N. Ahuja, "Robust multi-object tracking via cross-domain contextual information for sports video analysis," *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, 2012, pp. 985-988.
28. B. Sun, Z. Liu, Y. Sun, F. Su, L. Cao, and H. Zhang, "Multiple Objects Tracking and Identification Based on Sparse Representation in Surveillance Video," *Multimedia Big Data (BigMM), 2015 IEEE International Conference on*, Beijing, 2015, pp. 268-271.
29. S. Akkaladevi, M. Ankerl, C. Heindl, and A. Pichler, "Tracking multiple rigid symmetric and non-symmetric objects in real-time using depth data," *2016 IEEE International Conference on Robotics and Automation (ICRA)*, Stockholm, 2016, pp. 5644-5649.
30. T. Karpagavalli, and S. Appavu alias Balamurugan, "Multiple human tracking using 3D ellipsoidal model and SIR-Particle Filter technique," *Recent Trends in Information Technology (ICRTIT), 2014 International Conference on*, Chennai, 2014, pp. 1-5.
31. S. I. Yu, Y. Yang, and A. Hauptmann, "Harry Potter's Marauder's Map: Localizing and Tracking Multiple Persons-of-Interest by Nonnegative Discretization," *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, Portland, OR, 2013, pp. 3714-3720.
32. S. Rahimi, A. Aghagolzadeh, and H. Seyedarabi, "Three camera-based human tracking using weighted color and cellular LBP histograms in a particle filter framework," *2013 21st Iranian Conference on Electrical Engineering (ICEE)*, Mashhad, 2013, pp. 1-6.
33. Y. Yun, I. Y. H. Gu, and H. Aghajan, "Maximum-likelihood object tracking from multi-view video by combining homography and epipolar constraints," *Distributed Smart Cameras (ICDSC), 2012 Sixth International Conference on*, Hong Kong, 2012, pp. 1-6.
34. F. Vandewiele, F. Bousetouane, and C. Motamed, "Occlusion management strategies for pedestrians tracking across fisheye camera networks," *Distributed Smart Cameras (ICDSC), 2013 Seventh International Conference on*, Palm Springs, CA, 2013, pp. 1-6.
35. M. M. N. Ali, M. Abdullah-Al-Wadud, S. Lee ,"Multiple object tracking with partial occlusion handling using salient feature points" Information Sciences ,2014, pp. 448–465.

36. S. J. You, P. H. Truong, S. H. Ji, S. M. Lee, C. E. Lee and Y. J. Cho, "A cooperative multi-camera system for tracking a fast moving object," *Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), 2014 IEEE 4th Annual International Conf. on*, Hong Kong, 2014, pp. 141-145.
37. J. Fan *et al*., "Distributed multi-camera object tracking with Bayesian Inference," *2011 IEEE International Symposium of Circuits and Systems (ISCAS)*, Rio de Janeiro, 2011, pp. 357-360.
38. L. Zhang, J. Sturm, D. Cremers, and D. Lee, "Real-time human motion tracking using multiple depth cameras," *2012 IEEE/RSJ International Conf. on Intelligent Robots and Systems*, Vilamoura, 2012, pp. 2389-2395.
39. B. Kwolek, "Multi Camera-Based Person Tracking Using Region Covariance and Homography Constraint," *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*, Boston, MA, 2010, pp. 294-299.
40. A. Y. Yang, S. Maji, C. M. Christoudias, T. Darrell, J. Malik, and S. S. Sastry, "Multiple-view object recognition in band-limited distributed camera networks," *Distributed Smart Cameras, 2009. ICDSC 2009. Third ACM/IEEE International Conference on*, Como, 2009, pp. 1-8.
41. H. Li, and M. Flierl, "Sift-based multi-view cooperative tracking for soccer video," *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, 2012, pp. 1001-1004.
42. H. Li, Y. Li, and F. Porikli. DeepTrack: Learning discriminative feature representations by convolutional neural networks for visual tracking. In BMVC, 2014.
43. H. Fu, C. Wang, D. Tao and M. J. Black, "Occlusion Boundary Detection via Deep Exploration of Context," 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, 2016, pp. 241-250.
44. A. Toshev and C. Szegedy. *Deeppose: Human pose estimation via deep neural networks*. In CVPR, 2014..
45. A. Krizhevsky, I. Sutskever, and G. E. Hinton. *Imagenet classification with deep convolutional neural network*s. In NIPS, 2012.
46. J. Long, E. Shelhamer, and T. Darrell. *Fully convolutional networks for semantic segmentation*. In CVPR, 2015.
47. S. Hong, T. You, S. Kwak, and B. Han. *Online tracking by learning discriminative saliency map with convolutional neural network*. In ICML, 2015.
48. J. Fan, W. Xu, Y. Wu, and Y. Gong. *Human tracking using convolutional neural networks*. IEEE Trans. Neural Net- works, 21(10):1610–1623, 2010.
49. N.Wang, S. Li, A. Gupta, and D.-Y. Yeung. Transferring rich feature hierarchies for robust visual tracking. arXiv preprint arXiv:1501.04587, 2015.
50. R. Girshick, J. Donahue, T. Darrell, and J. Malik. *Rich feature hierarchies for accurate object detection and semantic segmentation*. In CVPR, 2014.
51. P. Babaei, and M. Fathy, "Abnormality Detection and Traffic Flow Measurement Using a Hybrid Scheme of SIFT in Distributed Multi Camera Intersection Monitoring," *2011 7th Iranian Conference on Machine Vision and Image Processing*, Tehran, 2011, pp. 1-5.
52. R. Feris, A. Datta, S. Pankanti and M. T. Sun, "Boosting object detection performance in crowded surveillance videos," Applications of Computer Vision (WACV), 2013 IEEE Workshop on, Tampa, FL, 2013, pp. 427-432.
53. P. Padeleris, X. Zabulis, and A. A. Argyros, "Multicamera tracking of multiple humans based on colored visual hulls," *2013 IEEE 18th Conference on Emerging Technologies & Factory Automation (ETFA)*, Cagliari, 2013, pp. 1-8.
54. Z. Liu, J. Huang; J. Han, S. Bu, J. Lv, "Human Motion Tracking by Multiple RGBD Cameras," in IEEE Transactions on Circuits and Systems for Video Technology , vol.PP, no.99, pp.1-1
55. M. Sigalas, M. Pateraki, and P. Trahanias, "Robust articulated upper body pose tracking under severe occlusions," *2014 IEEE/RSJ International Conf. on Intelligent Robots and Systems*, Chicago, IL, 2014, pp. 4104-4111.
56. X. Chen, Q. Meng, "Self-Learning Vehicle Detection and Tracking from UAVs," 2016, *International Journal of Mechanical Engineering and Robotics Research* Vol. 5, No. 2, April 2016, pp. 149-155.
57. Y. Cai, H. Wang, X. Sun, and L. Chen, "Visual Vehicle Tracking Based on Deep Representation andSemisupervised Learning," *Hindawi Journal of Sensors*Volume 2017, Article ID 6471250, pp. 1-6.

AUTHORS PROFILE

Dr. S. Sri Harsha working as Associate professor in Department of CSE, K L University having 10 years of experience in teaching and research, published good research papers in International journals and also published a patent in Vehicle Detection, Tracking & Classification. Research areas of interest are Digital Image Processing, Computer Networking Standards and Machine Learning.

Harika Simhadri is currently working as consulting Infrastructure Engineer in United States and has an insight in the development and operations of the software industry. The author has master's in Digital Sciences from Kent State University and has six years of software industry experience having keen research interest in the areas of Data Sciences and Image Processing.

Dr. Raghu Katragadda working as Sr. Assistant Professor in the department of business management at V.R. Siddhartha Engineering college, Vijayawada. Having over 9 years of industry as consultant and 10years of teaching experience has keen interest in the areas of data analytics and financial analytics.

Dr. K.V.Prasad working as Associate professor in Department of CSE, K L University having 14 years of teaching experience and 2 years of Industrial Exprerience. Reseach areas of intrest are Data Minining, Information Security,Machine Learning and Digital Image Processing. Published a petent on Internt of Things.