Dear Commissioner, please fix these: A scalable system for inspecting road infrastructure

Raghava Modhugu*, Ranjith Reddy*, and C.V. Jawahar

durga.nagendra@students.iiit.ac.in, ranjithreddy1061995@gmail.com, jawahar@iiit.ac.in
CVIT, IIIT-H

Abstract. Inspecting and assessing the quality of traffic infrastructure (such as the state of the signboards or road markings) is challenging for humans due to (i) the massive length of roads that countries will have and (ii) the regular frequency at which this needs to be done. In this paper, we demonstrate a scalable system that uses computer vision for automatic inspection of road infrastructure from a simple video captured from a moving vehicle. We validated our method on 1500 KMs of roads captured in and around the city of Hyderabad, India. Qualitative and quantitative results demonstrate the feasibility, scalability and effectiveness of our solution.

Keywords: Road Infrastructure · Scalable Audit · Indian Roads

1 Introduction and Related Work

Ever increasing traffic activity makes the regular audit and maintenance of road infrastructure extremely critical for safety. However, the scale at which this has to be performed in India is massive with thousands of kilometres of highways and rural roads in every state. In general road infrastructure maintenance and diligence are determined by financial constraints, and associated social factors. Road infrastructure inspection is a relentless process. The existing methods primarily use manual verification,which is tedious, cost ineffective and not scalable. Gaining insights through analytics is also difficult because of the lack of a unified system that stores all the necessary information of road infrastructure. In order to overcome these limitations and establish a robust road infrastructure management plan, we believe that inspection system needs to be highly automated, cost effective and scalable. In this paper, we propose a system that uses computer vision to automatically detect the road infrastructure and geotag [1] them with relevant attributes that state the condition of the infrastructure. We also measure the information about the visibility range of infrastructure, which plays an important role in case of traffic signs.

* equal contribution.

[1] Geotagging is the process of adding geographical information to various media in the form of metadata.

Fig. 1: In the proposed system, we aim to detect and classify the quality of the traffic signs, street lights and road lane markings to create database with geo-spatial information.

There is an increased interest in the field of computer vision for autonomous navigation in the recent years [6]. Computer vision based autonomous navigation systems try to exploit the visual instructions such as traffic signs and lane markings primarily meant for human navigational clarity. The advances in the areas of semantic segmentation and object detection paved way to numerous real world applications. In this paper, we propose a system that uses these aspects of computer vision to automatically detect the road infrastructure and geotag them with relevant attributes and their visibility. The proposed system helps in dealing with the challenges mentioned earlier.Our contributions in this paper are as follows:

1. We propose a scalable road infrastructure inspection system for detection based on state of art object detection approaches.
2. We propose a framework to geotag the detected infrastructure for precise location to save maintenance time.
3. We propose a framework to identify the condition of the road infrastructure and their visibility.
4. We release a dataset of road scenes for quality assessment of the infrastructure

A wide variety of road infrastructure requires maintenance, however in this work we choose to inspect traffic signs, street lights and lane markings. This choice is primarily based on the point that these infrastructure is present widely across the road. Therefore, these are a good choice for testing the scalability of the system. Secondly, we want to include not only just the object type infrastructure like traffic signs but also the infrastructure that is continuous and spread across regions, so that we can leave a very strong precedent on the road infrastructure inspection.

On testing our proposed system over 1500 KMs of road in Hyderabad to geotag the existing infrastructure with relevant attributes which will help in identifying in the state of traffic signs, roads without street lights and proper

lane marking. Out of the total 8323 traffic signs geotagged, 3308 (nearly 40%) of them are either rusted or faded. We found that 8323 geotagged traffic signs with attribute information and also contains information about street light distribution and lane marking quality on over 30000 stretches of road, with each stretch being 50m long.

Related work Numerous techniques have been developed in various parts of the world for road infrastructure audit independently. These techniques vary from manual to automated solutions. Manual techniques include field surveys and examination of recorded videos [5], [11]. The automated solutions in general use computer vision based techniques for detection, and global positioning systems to obtain geo-spatial information of the automatically detected infrastructure [3], [8], [9], [2], [12], [13]. The closest work to the proposed system in this paper is by Sudhir *et al.* [13] that focuses on auditing condition of roads, Whereas our audit system inspects a variety of infrastructure. Jones *et al.* [5] proposed a manual method where the personnel produces a detailed field survey and collect the GPS location of the infrastructure. A semi automated approach with the details of mobile mapping system (MMS) which requires data collection with a vehicle equipped with a global positioning system, distance measuring instrument and inertial navigational system is discussed by Khattak *et al.* [3]. It was experimentally concluded that the manually collected information is much more accurate than what was proposed by Khattak *et al.* [3]. A mobile data collection system presented by Maerz *et al.*[11] involves a post processing workstation to go through the recorded video to detect and classify the objects of interest. Jeyapalan *et al.* [8] used a method to determine the three-dimensional location of roadside features that appear in multiple images. This method maps the infrastructure without any classification or distinction between the features. A similar method as ours to recognise traffic signs and track them to avoid counting multiple times and map them using the GPS signal is proposed in Wang *et al.* [9]. In our system, we also track detected signs to avoid redundant counts. However, our system uses YOLOv3 as opposed to traditional image processing methods for detection, and tracking is used to determine the GPS position of the traffic sign. Unlike Wang *et al.*[9], our system is not just limited to traffic signs in road infrastructure. The traffic signs are detected and identified for inspection in Gonzalez *et al.* [2], but unlike our method it does not deal with the condition of the traffic signs. A compound architecture is proposed by Segvic *et al.* [12] for integrating independently developed vision components with the use of GPS to locate the traffic infrastructure. However, this system is restricted to object detection whereas our system inspects lane markings on the road also.

2 Framework and model

In this section, we discuss the architecture of the proposed system. The key aspects of the system is detecting, geotagging and inspection (quality identification). An overview of the system pipeline can be seen in Fig. 2.

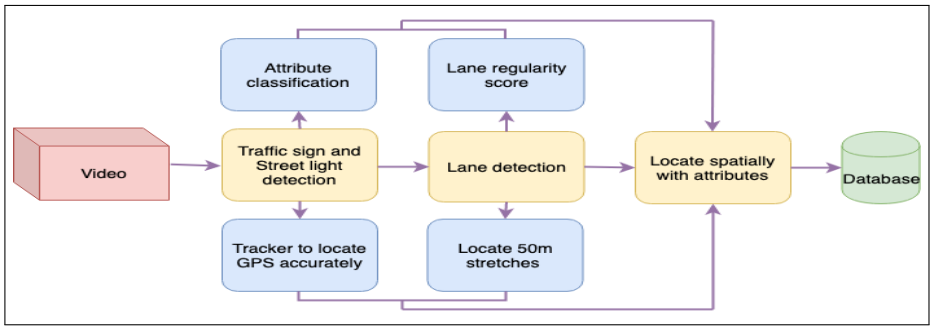

Fig. 2: The above figure illustrates the architecture of the proposed system for road infrastructure inspection. The traffic signs are detected using the YOLOv3. Once detected, these signs/lights are tracked to find the geographical location. In the next stage of the pipeline, SCNN is used to detect the lane markings. Using the detected lane markings by the SCNN, we generate a regularity score for every 50 meters stretch. Then using the obtained information of attributes and location we map the infrastructure to create a database.

Detection In this paper, the proposed system for road infrastructure inspections uses YOLOv3 [7] for detecting the traffic signs and street lights. Several object detection models can be trained to detect traffic signs and street lights, but YOLOv3 has many advantages such as object detection with global context, faster processing speed at test time. YOLOv3 can detect the objects at test time with 30 FPS on a Pascal Titan X and its single network evaluation makes it 1000× faster than R-CNN and 100× faster than fast R-CNN [7]. As the proposed system has to detect the traffic signs and street lights at real time speeds, YOLOv3 is a very natural choice for the proposed system.

As discussed in the introduction, we want to detect not only the traffic signs and street lights but also lane markings on the road. The proposed system detects lane markings using SCNN [15]. Unlike object detection, lane marking detection needs to tackle objects with strong structural prior but with less appearance clues. Lanes are continuous and might have been occluded by the vehicles in the traffic. It also requires precise prediction of the road curvature. In general, this can be done using probability maps on the image for lane markings [15]. SCNN [15] creates the probability maps for lane markings to identify them.

Geo tagging Geo-spatial information of the road infrastructure plays a very important role in decreasing the maintenance time by helping in easy identification. There are some methods proposed in the literature on locating the infrastructure with using GPS signal. The idea of mapping the infrastructure on a location where its bounding box area is maximum is proposed [14]. The other works that attempted to map the location of the infrastructure using photogrammetry are proposed in [1], [4]. In our work, we track the detected traffic signs or street light and map them to the location of the last frame it was detected while driving along the road. In case of lane marking detection, this method may

not be feasible since the lane is a continuous entity. We try to map the absence of a lane marking instead of the presence of it, which is of more interest for maintenance. More details on the criteria to find the absence will be provided in section 3.

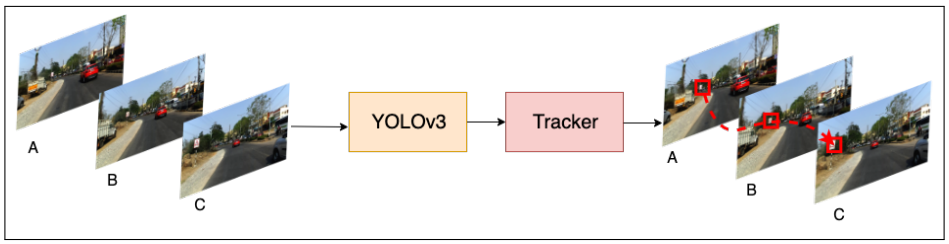

Fig. 3: The image illustrates the detection of infrastructure and tracking of it. We use tracker to identify the approximate GPS location and to find the visibility range of the traffic signs. Visibility range is calculated as the sum of distance between all the consecutive frames where the same traffic sign is detected using the GPS coordinates.

Inspection The key aspects of road infrastructure inspection involves assessing quality of the infrastructure. As the characteristics of quality are subjective to the type of infrastructure involved in assessment, different measures for quality assessment are required.

The traffic signs that are detected by the YOLOv3 are classified using a trained convolution neural network classifier with VGG16 architecture to find whether the detected sign is rusted, faded or normal. The attribute information extracted from the traffic sign is stored along with the geo-spatial information. As shown in Fig. 3, We also find the visibility range of each traffic sign as distance along the path traced by the vehicle from its first to final detection by tracking it. In case of street lights, we detect the streetlights using YOLOv3 to tag them geographically and identify all the stretches that donot have any street lights along the road. The lane markings are generally expected to be present continuously all along the road. The stretches of interest in this case are the ones without lane markings. We defined a metric that can take into account the percentage of pixels that have been identified as lane marking along a stretch of 50 meters of road to classify it accordingly.

Experimental setup We have captured the data of 1500 kilometers of real time data in and around the city of Hyderabad. Hyderabad region is deemed to be an appropriate region to collect the data to build and test the system because of various reasons such as high traffic activity, unstructured environment. The data is captured using a ZED camera which is mounted on a car. This captures 15 frames per second at a resolution of 1920×1080 pixels along with the GPS coordinates. The GPS is captured at baud rate of 4800 to synchronise it with the video using the time of capture. The data is captured in various areas to include

Fig. 4: Left: In anti clockwise direction: Intel's NUC, ZED camera, GPS unit and setup of all the components Right: Samples of images/frames from the collected data.

culture wise, maintenance wise, lighting wise and traffic conditions wise variety to suit the needs of the proposed inspection system.

3 Experiments and Results

Fig. 5: Qualitative results of traffic sign detection: The traffic signs are detected as objects using a trained YOLOv3. A variety of driving scenes where traffic signs are detected with variations in lighting and traffic conditions.

Traffic Signs We have used YOLOv3 for detection of traffic signs. A total of 2K frames are annotated manually which are used for training and validating the object detection model with a train-validation split of 80:20. The hyper parameters of the model are setup as suggested YOLOv3 [7] for custom object detection and trained for 6000 iterations and got an mAP of 0.58. We were able to achieve high recall of 0.91 on the unseen routes which is of key interest to the proposed system. The qualitative results in a wide variety of lighting and traffic conditions are illustrated in the Fig. 5.

The proposed system characterizes the detected signs to be rusted, faded or normal to aide the inspection. For this purpose we used VGG16 network. The pretrained VGG16 network is fine-tuned with a learning rate of 0.0001, L2 regularisation with decay of $1e-6$ using Adam on cross entropy loss. The data

required to train the classifier to characterise the traffic signs as rusted and faded are low in number. The collection of such data is very difficult and tedious. Therefore, we collected a few samples of rusted and faded traffic signs and used them for style transfer [10] on the traffic signs obtained from detected traffic signs using the trained YOLOv3 and also performed data augmentation techniques like data jittering, image transformation etc. to obtain around 7000 samples of rusted and faded traffic signs each. The accuracy obtained by the classifier is 91% on the validation data.

Fig. 6: Qualitative results of street lights detection: street lights along the road are detected using YOLOv3. The above figure shows a few scenes with street light detection.

Street Lights Street lights are also detected using YOLOv3. Street lights are manually annotated from the frames of the collected data to train the model with a train-validation split of 80:20. YOLOv3 [7] is trained for 4000 iterations to detect the street lights and obtained an mAP of 0.53. The qualitative results of detection are illustrated in Fig. 6. The street lights are mapped with GPS obtained and located on a map. This gives the information to gather the stretches where street lights are not installed.

Fig. 7: Qualitative results of road lane marking detection: The proposed system uses SCNN [15] to detect the lane marking along the road. The above figure illustrates lane marking detection in different conditions of traffic.

Road Lane Marking As discussed earlier, the method used to detect lane markings is very different from object detection methods in traffic signs and street lights images. The lane markings on the road are assumed to be a continuous entity. We choose pretrained SCNN proposed in [15] for lane marking detection in the system presented in this paper, majorly because of the reason that the network is trained with the data in an unstructured and heavy traffic environment which suits the proposed system. SCNN can also extrapolate the lane markings occluded by the vehicular traffic on the road based on structural priors obtained from the unoccluded parts as shown in the Fig. 7.

We are more interested in the stretches where the lane markings are absent. In order to find that out, we define a metric that uses the regularity of the lane markings for every 50 meters of distance as per the GPS data collected. Along with the distance, the system considers the time taken to cover that distance as well. There could be scenarios where video recording vehicle might have been stuck in traffic during the inspection, which we should be in a position to handle. Therefore we consider the metric as as summation of percentage of pixels present across all the frames to cover 50 meters. Lane regularity score (r) $= \dfrac{\sum_{i \in f} \dfrac{l_i}{n_i}}{d \cdot |f|}$.

l_i is number of pixels identified as lane marking in i^{th} frame, n_i is number of pixels in the i^{th} frame, d is length of the stretch, f is set of frames recorded to cover the stretch of distance

The number of pixel that belongs to lane marking are far less than total number of the pixels in an image. Therefore by definition lane regularity is a very small value generally ranging in the order of magnitude of 10^{-3} to 10^{-6} depending on the scene. So we normalised the lane regularity scores to bring all the values in the range of $[0, 1]$. Normalized lane regularity score $= \dfrac{r - r_{min}}{r_{max} - r_{min}}$. The r_{max} and r_{min} are maximum and minimum regularity scores determined experimentally.

Regions in Hyderabad	**Defective signs**	**Average distance between street lights**	**Lane marking that need attention**
Industrially active region	712	10.46 M	143.1 KM
Old city region	1000	19.58 M	169.85 KM
Suburban region	895	38.612 M	206.5 KM
Near by rural region	701	324.59 M	221.7 KM

Table 1: Quantitative results of road infrastructure in Hyderabad: The results shown above are obtained on 1500 KMs of test data that includes culturally, socially and financially diverse regions. With results shown in the above table we can conclude that the road maintenance is relatively better in the areas of industrially active areas than the other parts of the city, suburban areas and near by rural areas.

Quantitative Analysis After testing the system on 1500 KMs of data, we made the observations that are illustrated in the table 1. The median of average distance between two consecutive street lights in each route that was tested by the proposed system in Hyderabad is 24.84 meters and it is also observed that 49.43% of the total stretch of the road needs immediate attention to fix the lane markings. We also provided an analysis on the regularity of the lane markings by classifying all the 50 meter road stretches as fair, faded and unfair with the experimentally determined thresholds on normalised lane regularity score. The stretch with normalised lane regularity < 0.2 is considered an unfair road, if it is ≥ 0.2 and ≤ 0.5 is considered to be faded, it is > 0.5 then it is considered as a fair road with respect to the lane markings.

	Frames with detections	**True positives**	**False positives**	**False Negatives**	**Recall**	**Precision**
Traffic Signs	4999	80	43	6	0.93	0.65
Street lights	1408	112	4	47	0.70	0.96

Table 2: Quantitative results: The results shown in the above table are obtained on a 10 KMs of test road stretch. We manually annotated the road for street lights and traffic signs to find the recall and precision separately.

Discussions The quantitative and qualitative results illustrated in the previous section 3 strongly support our assumption that there is a lot of scope for improvement in road infrastructure maintenance in Hyderabad. It also supports our argument that the scalable and automated system is need of the hour and the proposed system is fit for that purpose. On observing the average consecutive street light separation of 24.84 meters, it can be concluded that the street light distribution is very sparse [2]. The results on lane marking with normalised regularity score of 0.2723 clearly state that the existing state of maintenance of lane markings is subpar and has to improve massively to meet the present day traffic. 40% of the total traffic signs are either rusted or faded, which again needs an massive upgrade in infrastructure. The observations are in concord with the general perception on the road infrastructure maintenance that the urban areas are well maintained when compared to the suburban and rural areas. In this work, we are able to detect and identify the quality of infrastructure that is present along the road at FPS of 2 .

4 Conclusion

We have presented an approach to build a road infrastructure audit system which can create a database of the targeted infrastructure with the geo-spatial

[2] The length of the stretch that is illuminated by a street light is highly dependent on its wattage and intensity. Even with a very optimistic consideration of 10 meter illumination by one street light, still the road is not optimal lit.

information along with relevant tags. Several future extensions to this work are possible such as recommending the missing infrastructure based on the scene and identifying the structures that creates occlusion on the road etc.

References

1. A. Gonzalez, L. M. Bergasa, and J. J. Yebes. Text detection and recognition on traffic panels from street-level imagery using visual appearance. *IEEE Transactions on Intelligent Transportation Systems*, 2014.
2. A. Gonzalez, M. A. Garrido, D. F. Llorca, M. Gavilan, J. P. Fernandez, P. F. Alcantarilla, I. Parra, F. Herranz, L. M. Bergasa, M. A. Sotelo, and P. Revenga de Toro. Automatic traffic signs and panels inspection system using computer vision. *IEEE Transactions on Intelligent Transportation Systems*, 2011.
3. Aemal J. Khattak, , Joseph E. Hummer, and Hassan A. Karimi. New and existing roadway inventory data acquisition methods. *Journal of Transportation and Statistics*, 2000.
4. Andrew Campbell, Alan Both, and Qian Sun. Detecting and mapping traffic signs from google street view images using deep learning and gis. *Computers Environment and Urban Systems*, 2019.
5. F.E. Jones. Gps-based sign inventory and inspection program. *IMSA Journal*, 2004.
6. Joel Janai, Fatma Güney, Aseem Behl, and Andreas Geiger. Computer vision for autonomous vehicles: Problems, datasets and state-of-the-art. *ArXiv*, 2017.
7. Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
8. Kandiah Jeyapalan. Mobile digital cameras for as-built surveys of roadside features. *Photogrammetric engineering and remote sensing*, 2004.
9. Kelvin Wang, Zhiqiong Hou, and Weiguo Gong. Automated road sign inventory system based on stereo vision and tracking. *Comp.-Aided Civil and Infrastruct. Engineering*, 2010.
10. Matthias Bethge Leon A. Gatys, Alexander S. Ecker. A neural algorithm of artistic style. *Nature Communications*, 2015.
11. Norbert H. Maerz and Steve McKenna. Surveyor: Mobile highway inventory and measurement system. *Transportation Research Record*, 1999.
12. S. Segvic, K. Brkić, Z. Kalafatić, V. Stanisavljević, M. Ševrović, D. Budimir, and I. Dadić. A computer vision assisted geoinformation inventory for traffic infrastructure. *IEEE Conference on Intelligent Transportation Systems*, 2010.
13. Sudhir Yarram, Girish Varma, and C. V. Jawahar. City-scale road audit system using deep learning. *International Conference on Intelligent Robots and Systems*, 2018.
14. Vahid Balali, Armin Ashouri Rad, and Mani Golparvar-Fard. Detection, classification, and mapping of u.s. traffic signs using google street view images for roadway inventory management. *Visualization in Engineering*, 2015.
15. Xingang Pan, Jianping Shi, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Spatial as deep: Spatial CNN for traffic scene understanding. *AAAI*, 2018.