Development of New Anpr Dataset for Automatic Number Plate Detection and Recognition in North of Iraq

Conference Paper · November 2019

DOI: 10.1109/UBMYK48245.2019.8965512

CITATIONS
4

READS
128

3 authors, including:

Naaman Omar
Duhok Polytechnic University
7 PUBLICATIONS **29** CITATIONS

SEE PROFILE

Abdulkadir Sengur
Firat University
221 PUBLICATIONS **4,700** CITATIONS

SEE PROFILE

Development of New Anpr Dataset for Automatic Number Plate Detection and Recognition in North of Iraq

Naaman Omar Yaseen
Department of Information Technology
Duhok Polytechnic University
Duhok, Iraq
naaman.omar@dpu.edu.krd

Salim Ganim Saeed Al-Ali
Department of Information Technology Management
Duhok Polytechnic University
Duhok, Iraq
salim.alali@dpu.edu.krd

Abdulkadir Sengur
Department of Electrical-Electronics Engineering
Firat University
Elazig, Turkey
ksengur@firat.edu.tr

***Abstract*- An automatic number plate detection (ANPD) and automatic number plate recognition (ANPR) systems are robust technologies that are used for detecting and recognizing the number plates of vehicles. In this paper, a new dataset, which is called North Iraq-Vehicle Images (NI-VI) of three provinces (Duhok, Erbil, and Sulaimani) for vehicle images, is presented. There are 1500 images in this dataset. They were gathered from real-time by using handled cameras to form a realistic dataset of the vehicle images. The main contribution of this work is the creation of a new dataset for license plate of vehicles in north Iraq with Arabic fonts in different and difficult conditions. The dataset includes three categories of images: rotated, scaled and translated images. The resolutions of images are 4288 x 2848 and 5184 x 3456. Moreover, some images created for bad weather conditions, such as snowy, dusty and low lighting. Some dirty plate images also considered in the dataset. The purpose of introducing this dataset is to provide and produce a realistic dataset for ANPD and as well as for ANPR systems.**

Keywords— Automatic number plate recognition (ANPR) datasets, Automatic number plate detection (ANPD) datasets, Plate number recognition datasets, and License number detection datasets.

I. INTRODUCTION

In the last decade, traffic control and traffic violation have become one of the most important issues in all countries. The flow of the traffic is controlled by many cameras that are spread around the streets inside and outside of cities [1,2]. As these cameras acquire instantaneous image of the vehicles, that are on the roads, intelligent software are needed to detect vehicles as well as their license plates. Generally, software needs training datasets for recognition of the license plates. Although of using these cameras, it is still difficult and challengeable to identify owner of the vehicle, who has violate the traffic, at the same moments using real-time application. It is very necessary to have such real-time dataset that will used to identify the vehicle owner, who has violate the traffic by driving with high speed more than the scheduled speed. Therefore, there is an urgent need to create a dataset for license plate of vehicles in each country, in order to investigate and solve the issue of ANPD and ANPR [3-7]. This dataset shall be used for three important purposes. First, it can be used for only vehicle number plate detection. Second, it can be used for only recognition the number plate, since There are labeled number plates in the dataset. Third, it can be used for both detection and recognition together, which means it can be used for ANPR systems.

In this paper, initially various license plate image datasets are reviewed [8-13]. More specifically, all these datasets are used for ANPR systems. These datasets contain images under different circumstances of weather conditions. Usually, these images are related to plate numbers that belong to one country. At the same time, in the other hand, some datasets are employed for automatic number plate detection (ANPD) systems [14-16]. All these datasets are used only for detection purpose (not recognition purpose). Moreover, there are some dataset that can be used for both detection and recognition the number plates in the images. The system of these kinds are called ANPR systems [17-22].

The rest of this paper is organized as follows, second section is about the datasets of ANPR literature reviews, in which set of datasets for vehicle images are reviewed and investigated all its properties. In the third section, is explained the ground truth of our proposed NI-VI dataset. Moreover, the fourth section is for comparison and discussion of these datasets. Finally, the conclusion of this paper is presented in the last fifth section.

II. DATASETS OF ANPR LITERATURE REVIEW

There are many datasets created for ANPR systems around the world. Each of these is concerned to plate numbers that belong to one country, and usually it's the researcher's country. In other words, there is some privacy, for each dataset, connected to the traffic rules of the country about the number plate features, such as, size, color, font, etc. In this section, different datasets of ANPR system are reviewed with more details about their images and features are presented.

978-1-7281-3992-0/19/$31.00 ©2019 IEEE

A. *The UFPR-ALPR Dataset [8]*

First, the UFPR-ALPR dataset is created for Brazil. It contains images that were fully annotated. These images were taken for 150 vehicles in real-world scenarios, and stored in 150 videos. All these videos with frame rate of 30 frames/second. All images, in these videos, were captured from moving (none stationary) vehicle, as well as were taken for moving vehicles, which means more realistic, during usual traffic in an urban environment. Be informed that this dataset consist of variety of vehicle images (cars, motorcycles, and trucks), it consists of 4500 images that have different conditions, such as backgrounds, lighting, plate positions, plate quality, vehicle types, and with various limitations for plates. Each license number plate has only 30 frame images. Finally, there are 1500 images are acquired by each of three different cameras: Huawei P9 Lite, iPhone 7 Plus, and GoPro Hero4 Silver. The PNG format is used to save images, with size of 1920×1080 pixels. The images of each camera are split into 3 parts: 900 of vehicle with gray license plates, 300 of vehicle with red plates, and 300 of motorcycles with gray ones. Examples about some images in the GTI dataset are shown in the Fig. 2.

The images have resolution of 360x256 pixels sequences with resolution of 64x64 pixels. This dataset is included images that captured from different point of views. Based on distance, the images are divided into two groups: middle, and far distance ranges. The middle range is also divided, based on the view of the vehicle, into three subgroups: left, center, and right. As a result of these division, four independent regions are created, each region involves 1,000 images of a certain view, the Fig. 2, below, shows some image examples of GTI dataset. The dataset is proposed different situations such as weather condition, and lighting.

Fig. 1. Image examples of UFPR-ALPR Dataset

The Brazilian license number plates are diverse in (size and color) based on the kind of vehicles and its category. Vehicle' LPs have 40cm _ 13cm of size, But motorcycles LPs have the size of 20cm _ 17cm. Also the color of cars LPs are variety according to vehicle type, for examples the LPs for private vehicles have gray color, while transportation vehicles, buses and taxis have red LPs. Also for other types of cars like older and official there are other LPs color used.

The images in the dataset divided in to three groups: 40% of the images for learning, 40% of the images for training and 20% of the images for validation. Each images has some explanatory notes in a file text, like by which camera the image taken, the position on the vehicle and types cars or motorcycle, due to the dataset contains both of them, also manufacturer, model and year, position of the LP and the position of its characters can annotated by the image.

B. *The GTI Dataset [9]*

The GTI dataset was created in highways of Spain (Madrid, Brussels and Turin). The images of this dataset were taken from a video sequences by a camera fixed on the front of the vehicle. In one hand, there are 3425 positive rear images, which contains the number plates within the vehicles, these images are extracted from variety locations of view. In the other hand, there are 3900 negative images, which contains the number plates without any vehicles' images), these images are taken from way sequences. There is a small number of images are used from Caltech and TU Graz-02 datasets, in order to approximate the total number of images into 4000 positive and 4000 negative images. Some of positive images contain vehicle rear completely, while others include about half of vehicle rear.

Fig. 2. Image examples of GTI Dataset, for different range of views

The GTI dataset is also proposed the percentages of 2000 of negative and positive images of different region of positions, 400 images are taken for each of these weather conditions: sunny, cloudy, and medium (neither very sunny nor cloudy). Another 400 images are taken for poor illumination such as down or dusk weather condition. For light raining weather condition, 200 image are used. The researchers include 100 images for bad resolution camera pictures. There are 50 images for industrial light, which are taken in the tunnels, are involved. Finally, there is 50 images out of the total 2000 images, the researchers did not mention about their weather conditions.

C. *The Markus Weber dataset [10]*

Markus Weber Cars dataset are taken by Markus Weber in California Institute of Technology's parking. It is not a very wide dataset since it includes only 126 images of resolution 896x592 pixels and all images are saved in JPG format. This dataset involves only images that taken from rear, and only for salon vehicles without including any track and bus vehicles. All images of this dataset acquired under the same conditions, which is only sunny days. Moreover, the dataset is not including any images that captured at night, low lighting, rain and shadow weathers. In addition to all these obstacles, there is no any tilt or rotation and clear translation in the dataset. The Fig.3. shows some image examples of Markus dataset.

Fig. 3. Image examples of Markus Weber Dataset

D. *The Baza-Slika Dataset [11]*

The Baza-Slika is a dataset of vehicle images created using Olympus Camedia C-2040Zoom digital camera. In this database, there are more than 500 images of the resolution used 640×460 for only rear views of cars. It includes three categories of vehicles: cars, trucks, and busses.

The images in this dataset are acquired from all over Croatia country through variety of lighting and weather conditions, for example there are seven folders of sunny, cloudy, sunshine, rainy, twilight and night light weather images. As well, the images are taken in diverse time of day, such as morning, after noon, evening and night with variant of qualities, brightness or contrasts, as shown as in the Fig. 4.

Fig. 4. Image examples of Baza-Slika Dataset

E. The SLVDS-iLPR Dataset [12]

This vehicle image dataset is created for Stop-Line Violation Detection System (SLVDS) dataset and is achieved for Indian Traffic Management system (ITMS). The images are collected using surveillance cameras of traffic monitoring of most metro cities in India.

The dataset consists of 4717 vehicle images, each image with resolution of 704x576 pixels. These images are taken from more than 30,000 snapshots of SLVDS. Some of these images are captured during the days, while others in the night and for different seasons of the year, which mean it covered different weather condition in India. Some examples about the vehicles' images of this dataset is shown below in the Fig. 5.

Fig. 5. Image examples of SLVDS-ITMS Dataset

F. The AOLP Dataset [13]

The Application Oriented license Plate (AOLP) database has been created by in the Artificial Vision lab, NTUST. The images in this dataset is classified into three categorization groups: Access Control (AC), traffic Law Enforcement (LE), and Road Patrol (RP).

The total number of images in this dataset include 2049 images with resolution of 320 x 240. In the first AC group, there are 681 images of moving, parking (stop), steady passing conditions. The distance between camera and number plate is equal 5m, the plate width is between 0.2 and 0.25 compare with the image's width. The second LE group has 757 images, which are taken by camera stand on road side. The images are for vehicles that violate traffic laws. The last RP group involves 611 images that captured from different points and variant distances using camera, which is handheld on a moving vehicle. The Fig. 6, below, shows some examples about the vehicles' images of this AOLP dataset.

Fig. 6. Image examples of AOLP Dataset

The images, in the AOLP dataset, are taken in different lighting conditions, such at night time, day time, outdoor and indoor. Also, the illumination of these images covers indoor, outdoor, daytime, nighttime, and various weather conditions.

III. GROUND TRUTH OF NI-VI DATASET

This section provides details and information related to vehicle images taken in different conditions, which are found in the present dataset. The aim is to improve ground truth of this dataset. Our new dataset is used for all different vehicle images of North of Iraq, so it is called NI-VI. This dataset with Arabic font (text and numbers) has been created to help researchers to apply their methods in automatic number plate detection and recognition systems to increase performance of these systems. Images in the proposed dataset are taken from real time by using two handled (unfixed) digital cameras from variety positions and angles. It involves images captured by using Canon 60D, EFS 18-55mm and Nikon DX, AF-S NIKKOR 18-105mm cameras of resolution 4288 x 2848 and 5184 x 3456 respectively.

Moreover, the (NI-VI) dataset comprises 1500 images taken from real time in different condition and variety weather situations such as day and night lighting with various backgrounds such as sunny, cloudy, snowy, foggy, dusty, and inside and outside cities. Furthermore, even some images of vehicles of dirty number plates are included in the dataset. Moreover, the images are captured in different times and places. Some of these images are taken under low or extra light source, other are taken under weak or strong sunlight. Our NI-VI dataset also involves images of different types of vehicles like Trucks, Buses and Salon with different colors for foreground and background colors. This differences is due that plate numbers are different. Some samples of images in NI-VI dataset are shown below in in Fig. 7.

Fig. 7. Image examples of NI-VI Dataset

In order to cover all aspects for 2D transformation in computer graphics, translation, scaling, and rotation are used in all different cases of the images in the dataset. Therefore, the images, within this dataset, are divided into three categories rotation, scale and translation. The fig. 7. Shows the three categories of all images in the NI-VI dataset. In the rotation category, the images are taken in both left and right direction. The angle slope between cars and cameras, inside the images, is ±20□. Totally, the rotation includes 400 images in four folders, such as 100 images for left angle with near distance, 100 images for left angle with far distance, 100 images for right angle with near distance and the last 100 images for right angle with far distance. For the scaling category, it involves 300 images based on the distance between the captured cars and the cameras, such as 100 images are taken from near distance, 100 images from mid and 100 images from far distance. In the third category for translation, the images are divided into two sub-categories: corners and sides. In one hand, 400 images are captured where the number plates are translated into the corners of the images, such as 100 images for left down corner, 100 images for left top, 100 images for right down and 100 images for right corner. In the other hand, 400 images are captured where the number plates are translated into the sides of the images, such as100 images for left side, 100 images for right, 100 images for top and the last 100 image for the bottom side. All categories and sub-categories of images inside our NI-VI dataset is shown below in the Fig. 8.

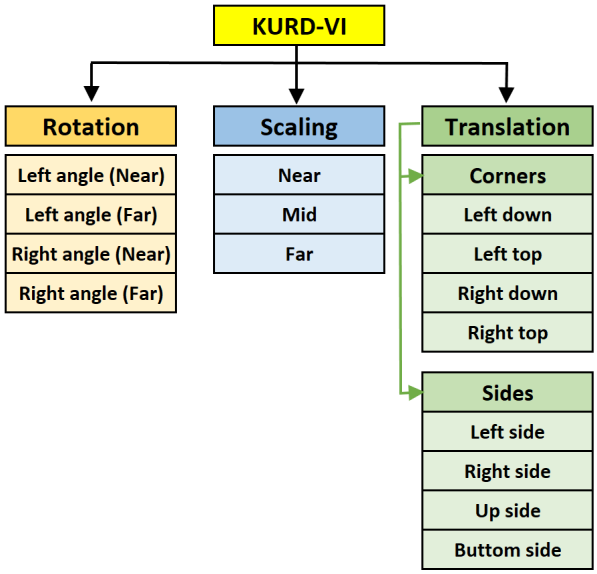

Fig. 8. Block diagram of NI-VI Dataset.

Furthermore, the presented dataset also involved 1500 cropped plate numbers from each whole image manually. Each of these images contains three main parts: (1) Top part involves license numbers. (2) The left of bottom part include provinces name. (3) The right of bottom part is used for country name. The names of these cropped license plate numbers are matched with names of whole images in our dataset. Some samples about these cropped images is shown below in Fig. 9.

Fig. 9. Cropped image examples of NI-VI Dataset

IV. Comparison and Discussion

The researchers, of the ANPD and ANPR approaches, are more likely interested to work with the number plates of their countries. Therefore, it is very urgent to design the system in global matter to detect and recognize any number plate regardless of its belonging country, since the cars can move among different countries. Thus, a dataset is created for north of Iraq for vehicles plates with Arabic text and numbers because there is no any dataset has been created for such kind of Arabic fonts.

A comparison among some reviewed datasets is achieved in this paper, as shown as in the Table I. This table compares these datasets based on some factors and attributes, such as country, year, and total number of images, resolution, and other conditions. Some of these datasets are old with bad resolutions, which there is a big lack of details inside the images. Nowadays, with the improvement happened in the cameras technologies would help to provide high resolution images but with more storage capacities. Furthermore, the weather conditions are not fully covered in some of these datasets, which mean that the dataset is far away from being as a realistic example of the real life. All the datasets that showed in the comparison table are representing the researchers' countries and as shown there is no any country with Arabic fonts. Thus, our suggested dataset is covering the lack of Arabic text and numbers in the vehicle images.

V. Conclusion And Future Works

In this paper, a new dataset (NI-VI) of north of Iraq vehicle image was systematically presented. All 1500 images in this dataset was gathered from real time by using handled cameras. The dataset includes three categories of images: rotation, scale and translation of resolution 4288 x 2848 and 5184 x 3456. The purpose of introducing dataset is to provide data for testing ANPD and ANPR algorithms by researchers to increase methods performance.

The NI-VI dataset was systematically presented in this paper. The importance of this dataset, it is required for ANPD and ANPR approaches and especially for north of Iraq. This dataset would help and assist the researches in these approaches. All works in the ANPD and ANPR approaches, in north of Iraq, should to be tested and experimented by some dataset that contains standardized images of vehicles and related to north

Iraq location. The limitation in this work is that this dataset related to vehicle license plates of only north Iraq, but it can be improved easily in the future work to include whole Iraq country.

TABLE I. COMPARISON OF DATASETS FOR NUMBER PLATE OF VEHICLES

Dataset Attributes					
Name	**Country**	**Year**	**Images**	**Res.**	**Conditions**
UFPR-ALPR	Brazil	2018	4500	1920 x 1080	Different backgrounds, lighting conditions, rear number plate positions, and cars type.
SLVDS-iLPR	India	2014	4717	704 x 576	Day morning, evening, night, sunny, rainy, cloudy, fog, shadow, low illumination, blurriness, various tilt angles and distances.
AOLP	Taiwan	2013	2049	320 x 240	Different lighting, day, night, indoor and outdoor illuminations.
GTI	Spain	2012	3425	360 x 256	Video sequences, sunny, cloudy, poor illumination, light rain, artificial lights.
Markus Weber	USA	2003	126	896 x 592	Parking cars, sunny days, and rear center of small cars.
Baza-Slika	Croatia	2001	500	640 x 460	Sunny, cloudy, sunshine, rainy, twilight and night light, rear view left and right rotation, near scaling.
Our dataset (NI-VI)	**North of Iraq**	**2019**	**1500**	**4288 x 2848 and 5184 x 3456**	**Different conditions, night light, day, sunny, snow, fogy, dirty, shadow, cloudy, rainy, front, rear, rotation, scale, translation, unfixed distance, two-handle camera, different angles.**

REFERENCES

[1] K. Raghunandan et al., "Riesz fractional based model for enhancing license plate detection and recognition," IEEE Trans. Circuits Syst. Video Technol., vol. 28, no. 9, pp. 2276–2288, 2017.

[2] L. Hu and Q. Ni, "IoT-driven automated object detection algorithm for urban surveillance systems in smart cities," IEEE Internet Things J., vol. 5, no. 2, pp. 747–754, 2017.

[3] N. Saleem, H. Muazzam, H. Tahir, and U. Farooq, "Automatic license plate recognition using extracted features," presented at the 2016 4th International Symposium on Computational and Business Intelligence (ISCBI), 2016, pp. 221–225.

[4] S. Du, M. Ibrahim, M. Shehata, and W. Badawy, "Automatic license plate recognition (ALPR): A state-of-the-art review," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 2, pp. 311–325, 2012.

[5] S. Du, M. Ibrahim, M. Shehata, and W. Badawy, "Automatic license plate recognition (ALPR): A state-of-the-art review," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 2, pp. 311–325, 2012.

[6] M. S. Al-Shemarry, Y. Li, and S. Abdulla, "An efficient texture descriptor for the detection of license plates from vehicle images in difficult conditions," *IEEE Trans. Intell. Transp. Syst.*, 2019.

[7] O. I. Al-Sanjary, A. A. Ahmed, and G. Sulong, "Development of a video tampering dataset for forensic investigation," *Forensic Sci. Int.*, vol. 266, pp. 565–572, 2016.

[8] R. Laroca *et al.*, "A robust real-time automatic license plate recognition based on the YOLO detector," presented at the 2018 International Joint Conference on Neural Networks (IJCNN), 2018, pp. 1–10.

[9] J. Arróspide, L. Salgado, and M. Nieto, "Video analysis-based vehicle detection and tracking using an MCMC sampling framework," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, p. 2, 2012.

[10] M. Oliveira and V. Santos, "Automatic detection of cars in real roads using haar-like features," Dep. Mech. Eng. Univ. Aveiro, vol. 3810, 2008.

[11] "Projekt 'License Plates.'" [Online]. Available: http://www.zemris.fer.hr/projects/LicensePlates/english/. [Accessed: 17-Jul-2019].

[12] S. Saha, S. Basu, and M. Nasipuri, "iLPR: an Indian license plate recognition system," *Multimed. Tools Appl.*, vol. 74, no. 23, pp. 10621–10656, 2015.

[13] G.-S. Hsu, J.-C. Chen, and Y.-Z. Chung, "Application-oriented license plate recognition," *IEEE Trans. Veh. Technol.*, vol. 62, no. 2, pp. 552–561, 2012.

[14] L. Xie, T. Ahmad, L. Jin, Y. Liu, and S. Zhang, "A new CNN-based method for multi-directional car license plate detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 2, pp. 507–517, 2018.

[15] S. He, Y. Yuan, C. Fu, X. Hu, and Y. Zhao, "Robust license plate detection using profile-based filter," presented at the 2018 Tenth International Conference on Advanced Computational Intelligence (ICACI), 2018, pp. 794–800.

[16] S. Yu, B. Li, Q. Zhang, C. Liu, and M. Q.-H. Meng, "A novel license plate location method based on wavelet transform and EMD analysis," *Pattern Recognit.*, vol. 48, no. 1, pp. 114–125, 2015.

[17] M. S. Al-Shemarry, Y. Li, and S. Abdulla, "Ensemble of adaboost cascades of 3L-LBPs classifiers for license plates detection with low quality images," *Expert Syst. Appl.*, vol. 92, pp. 216–235, 2018.

[18] K. Deb, H.-U. Chae, and K.-H. Jo, "Vehicle License Plate Detection Method Based on Sliding Concentric Windows and Histogram.," *JCP*, vol. 4, no. 8, pp. 771–777, 2009.

[19] S. M. Silva and C. R. Jung, "Real-time brazilian license plate detection and recognition using deep convolutional neural networks," presented at the 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), 2017, pp. 55–62.

[20] S. Kaur and S. Kaur, "An efficient approach for number plate extraction from vehicles image under image processing," *Int. J. Comput. Sci. Inf. Technol.*, vol. 5, no. 3, pp. 2954–2959, 2014.

[21] M. R. Asif, Q. Chun, S. Hussain, M. S. Fareed, and S. Khan, "Multinational vehicle license plate detection in complex backgrounds," *J. Vis. Commun. Image Represent.*, vol. 46, pp. 176–186, 2017.

[22] N. More and B. Tidke, "License plate identification using artificial neural network and wavelet transformed feature selection," presented at the 2015 International Conference on Pervasive Computing (ICPC), 2015, pp. 1–5.